Founded by G. Stanley Hall in 1891, reorganized by Carl Murchison in 1925

JOURNAL OF GENETIC PSYCHOLOGY

Child Behavior, Animal Behavior, and Comparative Psychology

EDITED BY
CARL MURCHISON

JOHN E. ANDERSON
University of Minnesota

S. K. BOSE
University of Calcutta

CHARLOTTE BÜHLER
University of Southern California

CYRIL BURT
University of London

LEONARD CARMICHAEL
Smithsonian Institution

SANTE DE SANCTIS
R. Università di Roma

ARNOLD GESELL
Yale University

WILLIAM HEALY
Judge Baker Foundation, Boston

ARTHUR T. JERSILD
Teachers College, Columbia University

BUFORD JOHNSON
The Johns Hopkins University

HAROLD E. JONES
University of California

TRUMAN L. KELLEY
Harvard University

H. P. MAITI
Institute of Child Development, Ahmedabad, India

TORAO OBONAI
Tokyo University, Japan

YOSHIKAZU OHWAKI
Tohoku University, Japan

HENRI PIÉRON
Université de Paris

ROBERT R. SEARS
Stanford University

GEORGE D. STODDARD
University of Illinois

C. J. WARDEN
Columbia University

JOHN B. WATSON
New York City

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 93

1958

Copyright, 1958, by The Journal Press
Published Quarterly by The Journal Press
Provincetown, Massachusetts, U. S. A.

INDEX OF AUTHORS

Adlerstein, M. 167
Alexander, Irving E. 167
Bartlett, Claude J. 153
Blakely, W. Paul 291
Boardman, William K. 185
Brown, W. Lynn 73, 287
Burke, Henry R. 199
Davis, R. T. 37
Dixon, J. C. 79
Feinberg, Mortimer R. 133
Goldstone, Sanford 185
Gordon, Anitra 303
Goss, Albert E. 191
Grodsky, M. A. 37
Hopper, Harold E. 3
Horrocks, John E. 153
James, W. T. 179
Jones, Mary Cover 87
Levinson, Boris M. 237, 277
Lhamon, William T. 185
McDowell, A. A. 37, 63
McDowell, R. A. 73
Milton, Alexander 249
Moylan, Marie C. 191
Nash, Harvey 149
Pinneau, Samuel R. 3, 249
Plant, Walter T. 53
Pronko, N. H. 113
Schmidt, Robert 133
Shaklee, Alfred B. 59
Smith, Max 133
Steele, J. P. 37
Thayer, Lee O. 113
Thurlow, Willard 303
Wang, James D. 119
Warren, J. M. 33, 229
White, Robert K. 287
Winthrop, Henry 309
Zuk, G. H. 15

Edn. Research
Dated 3.9.70
J-315

TABLE OF CONTENTS

The relationship between incidence of specific gastro-intestinal reactions of the infant and psychological characteristics of the mother 3
BY SAMUEL R. PINNEAU AND HAROLD E. HOPPER

Sex-appropriate behavior in adolescence 15
BY G. H. ZUK

Effects of satiation on food preferences in monkeys 33
BY J. M. WARRREN

The performance of X-ray irradiated and nonirradiated rhesus monkeys before, during, and following chronic barbiturate sedation 37
BY R. T. DAVIS, A. A. MCDOWELL, M. A. GRODSKY, AND J. P. STEELE

Sex, intelligence, and sorority or fraternity membership and changes in ethnocentrism over a two-year period 53
BY WALTER T. PLANT

Chlorpromazine and habit reversal 59
BY ALFRED B. SHAKLEE

Comparisons of distractibility in irradiated and nonirradiated monkeys . . 63
BY ARNOLD A. MCDOWELL

Facilitative effects of irradiation on performance of monkeys on discrimination problems with reduced stimulus cues 73
BY A. A. MCDOWELL AND W. LYNN BROWN

Reactions of superior and feebleminded children to an illusion . . . 79
BY J. C. DIXON

A study of socialization patterns at the high school level 87
BY MARY COVER JONES

Some psychological factors in the reading of fiction 113
BY LEE O. THAYER AND N. H. PRONKO

The relationship between children's play interests and their mental ability 119
BY JAMES D. WANG

An analysis of expressions used by adolescents at varying economic levels to describe accepted and rejected peers 133
BY MORTIMER R. FEINBERG, MAX SMITH, AND ROBERT SCHMIDT

Stereotyped associations to schematic faces 149
BY HARVEY NASH

A study of the needs status of adolescents from broken homes 153
BY CLAUDE J. BARTLETT AND JOHN E. HORROCKS

BOOKS 161

BOOKS RECENTLY PUBLISHED 161

Affective responses to the concept of death in a population of children and early adolescents 167
By Irving E. Alexander and Arthur M. Adlerstein

Conditioned responses in the opossum 179
By W. T. James

Kinesthetic cues in the development of time concepts 185
By Sanford Goldstone, William K. Boardman, and William T. Lhamon

Conceptual block-sorting as a function of type and degree of mastery of discriminative verbal responses 191
By Albert E. Goss and Marie C. Moylan

Raven's progressive matrices: A review and critical evaluation 199
By Henry R. Burke

The development of paw preferences in cats and monkeys 229
By J. M. Warren

Reevaluation of the Revised Stanford-Binet Scale, Form *L* Vocabulary as a test of intelligence for the kindergarten and primary school child . . 237
By Boris M. Levinson

The ecological veracity of the self-report 249
By Samuel R. Pinneau and Alexander Milton

Cultural pressure and *WAIS* scatter in a traditional Jewish setting . . . 277
By Boris M. Levinson

Preirradiation fatigue as a factor in the prevention of irradiation deaths in rats 287
By W. Lynn Brown and Robert K. White

A study of seventh grade children's reading of comic books as related to certain other variables 291
By W. Paul Blakely

SHORT ARTICLES AND NOTES 303

Substitution with interrupted tasks of differing valence (Anitra Gordon and Willard Thurlow) 303

BOOKS 307

BOOKS RECENTLY PUBLISHED 307

Packard, V. *The Hidden Persuaders.* (Reviewed by Henry Winthrop) 309

The Journal of Genetic Psychology, 1958, **93**, 3-13.

THE RELATIONSHIP BETWEEN INCIDENCE OF SPECIFIC GASTRO-INTESTINAL REACTIONS OF THE INFANT AND PSYCHOLOGICAL CHARACTERISTICS OF THE MOTHER*

Institute of Child Welfare, University of California, Berkeley, and Department of Psychology, University of Oregon

SAMUEL R. PINNEAU AND HAROLD E. HOPPER[1, 2]

A. INTRODUCTION

The effects on the child of the psychological characteristics of the mother have long been recognized. Within recent years there has been an increased stress on the importance of these characteristics in determining the reactions of the infant: Sullivan, for example, states that ". . . the tension of anxiety in the mothering one induces anxiety in the infant" (11, p. 41). Ribble maintains that children who do not obtain sufficient mothering develop gastro-intestinal disorders as shown, for example, in constipation and frequent vomiting (8, p. 631) and she states that the emotionally disturbed mother is as unable to mother a child as she is to secrete milk (8, p. 633). Montague (6), relying heavily on the works of Sontag (9, 10), states that emotional disturbances in the mother during pregnancy result in excessive regurgitation and other gastro-intestinal disorders in the infant. And Escalona (1) has suggested that there is a relationship between the emotional states of the adult and the feeding reactions of the infant.

The primary concern of this paper is the relationship between the mother's psychological characteristics and the incidence of gastro-intestinal disturbances in the infant. Of the above mentioned writings bearing specifically on this problem, Escalona's is based on the most extensive observation. Her observations, made at the Massachusetts Reformatory for Women in Framingham during the period from 1938 to 1940, were on a group of 50 to 60 babies of whom 70 per cent were less than one year old. Observations were made in four different departments of the institution which were separated

---

*Accepted for publication by Harold E. Jones of the Editorial Board, and received in the Editorial Office on April 22, 1957.

[1] The data analyzed in this study were collected while the investigators were at the University of Oregon. A first report on the results was presented at the thirty-fourth meeting of the Western Psychological Association (3).

[2] At the time of the final analysis of the data the first author was a Public Health Service Research Fellow of the National Institute of Mental Health.

spatially by as much as a half mile. "Her contact with the children was such that two important obstacles ordinarily encountered by clinicians were circumvented in that the children were under direct observation for 24 hours daily over long periods of time and that the information was obtained first hand instead of through informants of varying degrees of reliability" (1, p. 76-77).[3]

According to Escalona's report:

> It held true for the entire group that the eating situation was a very sensitive indicator of the emotional atmosphere prevailing in the institution at large. On days which had a special emotional significance for the students, such as Christmas, holidays or "parole day," this was especially marked. The incidence of meal time tantrums, refusal of food, vomiting and other eating disturbances among babies was always much higher on parole days although the routine was exactly the same (1, p. 80).

And, according to her conclusion: "When infants and young children are brought into close contact with an adult, they perceive the emotional status of the adult and respond to it in a consistent manner" (1, p. 80).

However, she also states: "The data reported are not to be regarded as the outcome of a systematic study. On the contrary, they are more or less isolated facts which came to attention during the stay at Framingham" (1, p. 76); and "It is obvious that it would be scientifically unsound to draw conclusions or make generalizations on the basis of data such as these" (1, p. 80).

Findings such as Escalona's suggest that we might expect to find a relationship between the emotional states of the mother and digestive disturbances in the infant as reflected in frequency of regurgitation. An additional hypothesis is suggested by Ribble's contention that the more maladjusted mothers are less able to mother their children (8, p. 633) and that these less stimulated children swallow more air (8, p. 631), namely, that burping might be expected to occur more often among these infants. These are the main hypotheses of the present investigation.

B. Subjects and Procedure

In order to obtain subjects for the present study, letters explaining something of its nature were sent to 400 parents of infants in the Eugene-Springfield-Coburg area of Oregon. A favorable response was obtained from 88 of the parents. Of these, 55 reported some occurrence of regurgitation in their infants. These mothers and their infants were chosen as subjects. Subsequently, for purposes of another study (4), the subjects were randomly assigned to a control or experimental group and matched for age; this procedure yielded 21 pairs. Beginning with the second week of the study the experimental group received additional "mothering" in order to determine if increased stimulation affected frequency of regurgitation; the results did not support this hypothesis. Thirty-eight of the mothers in the total sample took the *MMPI* and remained in the study for the full five weeks. During this time they tallied the occurrence of regurgitations, feedings, burps, and new vocalizations. Generally the mothers were above the average in terms of educational level and socio-economic status. [The sampling and experimental procedures are described in detail in an earlier publication (4).]

---

[3] The quotations in this section from S. Escalona, *Amer. J. Orthopsychiat.*, 1945, **15**, 76-80, are made with the permission of the American Orthopsychiatric Association, Inc.

The *MMPI* was chosen as the appropriate measuring instrument in this study because it provides the most widely accepted and objective measures of anxiety and emotional disturbances. The obtained protocols were scored on four measures of anxiety: Taylor's Manifest Anxiety Scale (12), Welsh's Scale of Judged Anxiety, his "*A*" factor, and his Anxiety Index (14). The *MMPI*'s were also scored on the usual scales and on Winne's Neuroticism Scale (16).

The psychological characteristics of the mother as measured by the *MMPI* were correlated with the incidence of their infants' regurgitation and burping as recorded over the five-week period.

C. RESULTS

1. *Consistency of Gastro-Intestinal Reactions*

In the earlier report (4) correlations presented indicate a marked tendency for the subjects to maintain their same relative position in terms of frequency of regurgitation. (For adjacent weeks the correlations ranged between .82 and .98.) Because the subjects differed in age by as much as three months, the question was investigated as to whether heterogeneity with respect to age resulted in a spurious picture of stability: It was possible to find 22 of the subjects who had been observed for the same age period of 99-113 days. The correlation between frequency during the first and second week of this period was .94. Evidence that relative position was essentially the same as in the preceding groupings was obtained by correlating these subjects' frequencies for the age interval 106-113 days with frequency reported during the first week of observation at which time they varied in age by as much as 26 days ($r = .87$). An alternative check on this is the finding that for the 38 subjects in the present investigation, the correlation between age of beginning the study and average frequency over the five-week period was only .04. Thus during this period of infancy it appears that heterogeneity with respect to age contributes little if any to the consistency of relative standing on this variable.

Checks indicate similar consistency in relative standing in terms of incidence of burping. The correlation between the average frequency of burping during the first and third weeks of the experimental period for the combined experimental and control groups was .90. This consistency of relative standing suggests that the increased handling in the case of the experimental subjects did not materially effect the incidence of burping. This hypothesis is supported by a direct test; the change of frequency in the experimental group from the first to the third week of the experimental period was not significantly different from that found for the control group. For 20 of the pairs "t" equalled 1.58.[4] It appears that this insignificantly greater change of the experimental group can be explained in terms of the regression phenomena, since by chance it contained a larger proportion of the extreme cases (4).

2. *Incidence of Gastro-Intestinal Reactions in the Infant as Related to "Anxiety in the Mothering One"*

The relationship between frequency of regurgitation in the infants over the total five-week period of the study and the mothers' anxiety as measured by four *MMPI* scales are given in Figure 1 ($N = 38$). The relationships are not significant; however, it is interesting to note that the tendencies are in a direction opposite to that expected on the basis of the authors' works to which we have referred. The means and standard deviations of the anxiety scores are given in Table 1.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF ANXIETY MEASURES

	Mean	*SD*'s
Welsh's Scales:		
Anxiety Index	19.3	9.82
Judged Anxiety	5.6	4.00
A Factor	9.7	7.04
Taylor's Scale	14.3	9.19

---

[4] One of the 21 pairs originally established (4) had to be excluded in this comparison because one of the mothers failed to keep a complete record on the incidence of burps.

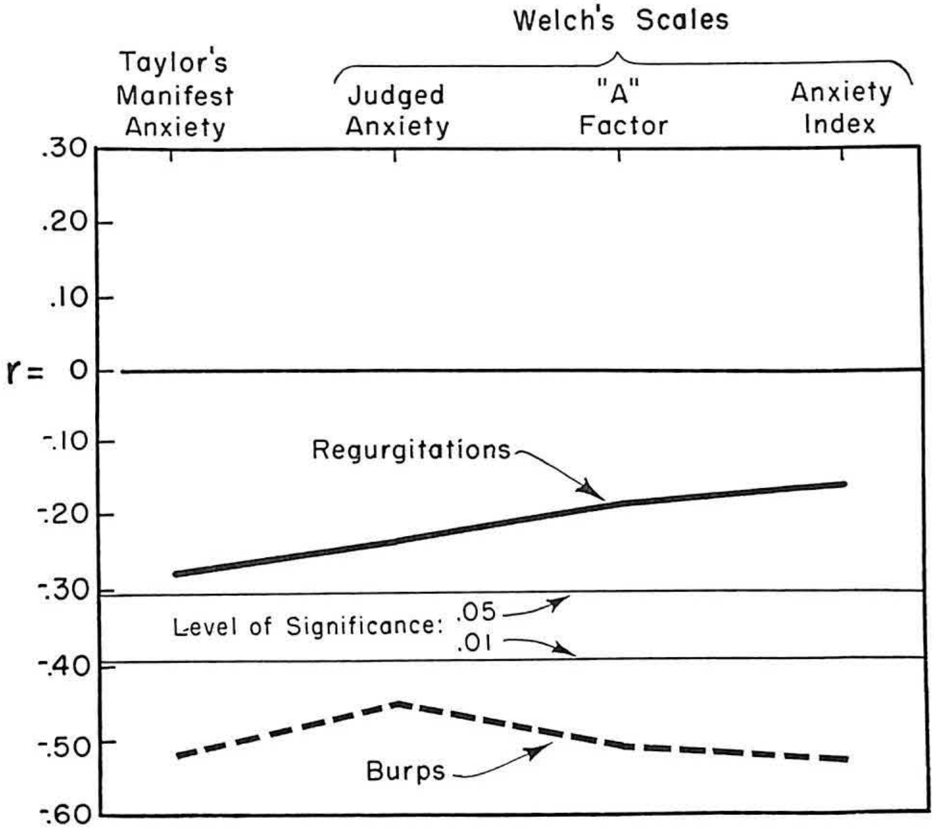

FIGURE 1
GASTRO-INTESTINAL REACTIONS OF INFANTS AS RELATED TO *MMPI* MEASURES OF MOTHERS' ANXIETY

Kristal and Pinneau (5) have reported that score on Taylor's Manifest Anxiety Scale is not sensitive to transitory anxiety and that the test-retest reliability coefficients over several months are quite high. If this is characteristic of anxiety scales, it might be maintained that their use will give a poor test of the hypothesis. However, if there is a high degree of individual consistency in frequency of regurgitation, this criticism would hardly seem applicable, that is, the anxiety reflected in frequency of regurgitation must also be quite stable. The relationships previously reported indicate a high degree of consistency; however, these were based on weekly averages. The question arises regarding stability of regurgitations from day to day. Correlations between frequencies for the combined control and experimental groups were determined between the first and second, and between the seventh and eighth days of the experimental period; the correlations of .88 and .75 give little support to the notion of marked daily fluctuations in the individual's relative standing in frequency of regurgitation. Thus it appears unlikely that frequency of regurgitation is fluctuating with transitory anxieties of the "mothering one." It would seem only if the relationship had been quite low that such a hypothesis would be tenable.

Figure 1 also gives the relationship between anxiety in the "mothering one" and the gastro-intestinal reaction, frequency of burping. These relationships are considerably higher than those found between anxiety and incidence of regurgitations, all of them reaching the one per cent level of significance. However, the results do not support the hypothesis set forth on the basis of Ribble's report, namely, that frequency of burping is positively related to anxiety in the "mothering one," since the obtained relationships were all negative. An alternative explanation is that the relationship can be accounted for in terms of closer and more active contact between non-anxious mothers and their children; i.e., the non-anxious mother engages in more activities with the child and the increased stimulation leads to a higher incidence of regurgitation and burping. However, this hypothesis does not appear warranted in the light of the finding that frequency of burping in the experimental group was not significantly affected by increased stimulation.

3. *Incidence of Gastro-Intestinal Reactions as Related to Other Psychological Characteristics of the Mother*

The mothers' *MMPI* protocols were scored on the standard scales and also on Winne's Neuroticism Scale. The means and standard deviations for these measures are given in Table 2. The correlations between these measures of the mothers' characteristics and the incidence of the infants' gastro-intestinal reactions of regurgitation and burping are given in Figures 2 and 3 respectively. The hypotheses, based on earlier studies, of positive relationships between these gastro-intestinal reactions and extent of the mothers' maladjustments, are not supported by these results. In fact, in those instances in which the correlations are significant, the directions of the relationships are opposite to that predicted from the earlier studies. Thus the converse hypothesis is again suggested, i.e., that closer and more active contact between the well-adjusted mothers and their children results in their children having higher frequency of "gastro-intestinal reactions"; but, as before, this interpretation does not appear warranted in view of the

TABLE 2
MEANS AND STANDARD DEVIATIONS OF MOTHERS' SCORES ON CONVENTIONAL SCALES OF THE *MMPI*

	Uncorrected scores		Scores Corrected with *K*	
	Mean	*SD*'s	Mean	*SD*'s
L	3.9	2.13		
F	2.6	1.88		
K	17.2	4.80		
Hs	4.9	4.43	13.8	4.12
D	20.4	3.70		
Hy	22.3	4.56		
Pd	13.2	4.13	20.1	3.70
Mf	38.6	4.67		
Pa	9.5	2.48		
Pt	10.7	6.97	28.0	5.04
Sc	8.0	6.42	25.2	5.08
Ma	13.9	3.93	17.4	3.50
Si	25.4	8.38		
Winne's Scale	5.3	3.52		

finding that increased stimulation does not significantly affect the incidence of burping or regurgitation (4).

D. DISCUSSION

In contrast to earlier studies of the relationship between infants' gastro-intestinal reactions and the emotional status of the "mothering ones," the present study used objective measures both for the gastro-intestinal reactions and for the psychological characteristics. Generally the results do not appear to be consistent with those obtained in the earlier studies since they suggest at least as high, if not a higher, incidence of gastro-intestinal disturbance (as shown by frequency of regurgitation and of burping) among the infants from the less anxious and/or better adjusted mothers as from the more anxious and/or poorer adjusted mothers.

The consistent tendency for frequency of regurgitation and of burping to be negatively related to measures of the mothers' tendencies towards maladjustment might be accounted for in a number of ways: The first way, already considered and rejected as unlikely, is that a closer and more active contact between the well adjusted mother and her child results in the increased incidence. An alternative possibility is that these gastro-intestinal reactions are maturity indices. As such, they would suggest less maturity in the case of the children of the better adjusted (more mature?) mothers. If so, it is conceivable, considering the generally positive relationship between the parents' characteristics and the characteristics of their children at ma-

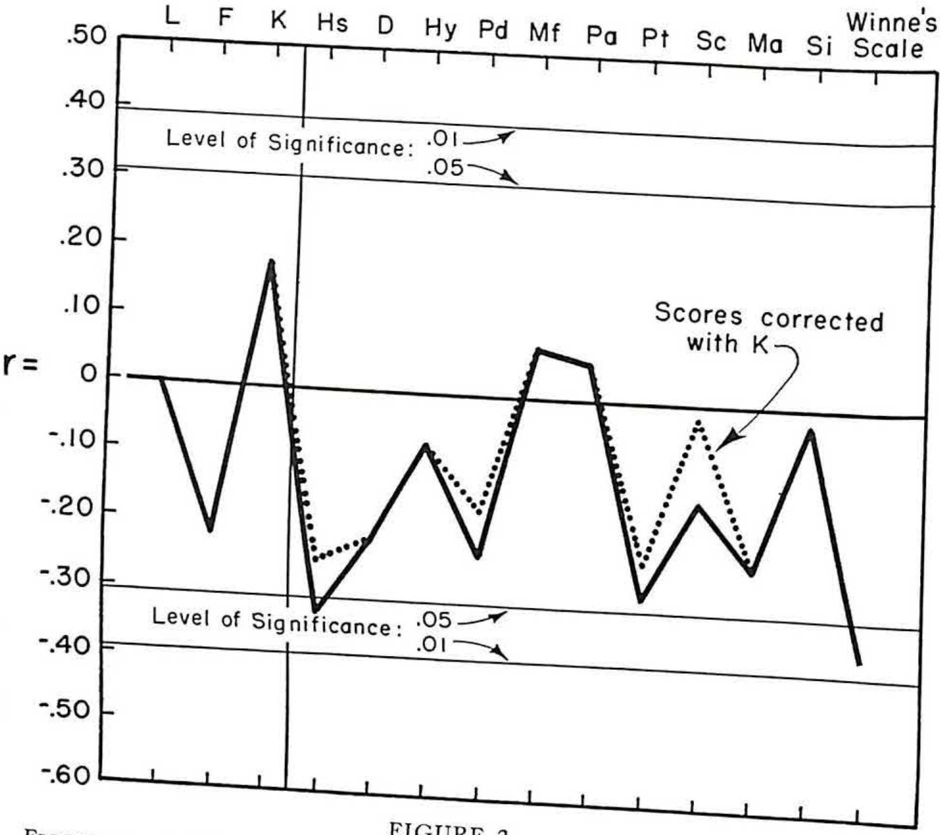

FIGURE 2
FREQUENCY OF REGURGITATION OF INFANTS AS RELATED TO MOTHERS' SCORES ON CONVENTIONAL *MMPI* SCALES

turity, that this is another instance in which slow maturing in early infancy tends to be associated with an end state of greater than average maturity. Another alternative is that the less adequately adjusted mothers are less accurate observers, that is, their observation of these reactions tend to be attenuated by personal problems and preoccupations. The high relationships between incidence reported in the original questionnaire and incidence tallied in the observational record (4) suggests that, if present, this tendency was not one of "carelessness" or of indifference toward the project, but a broader or more general trait. A fourth alternative explanation for the consistency of these results is in terms of overlap of scales and lack of independence of the traits; that is, if, by chance or otherwise, these gastrointestinal reactions are related to one of the mothers' characteristics measured

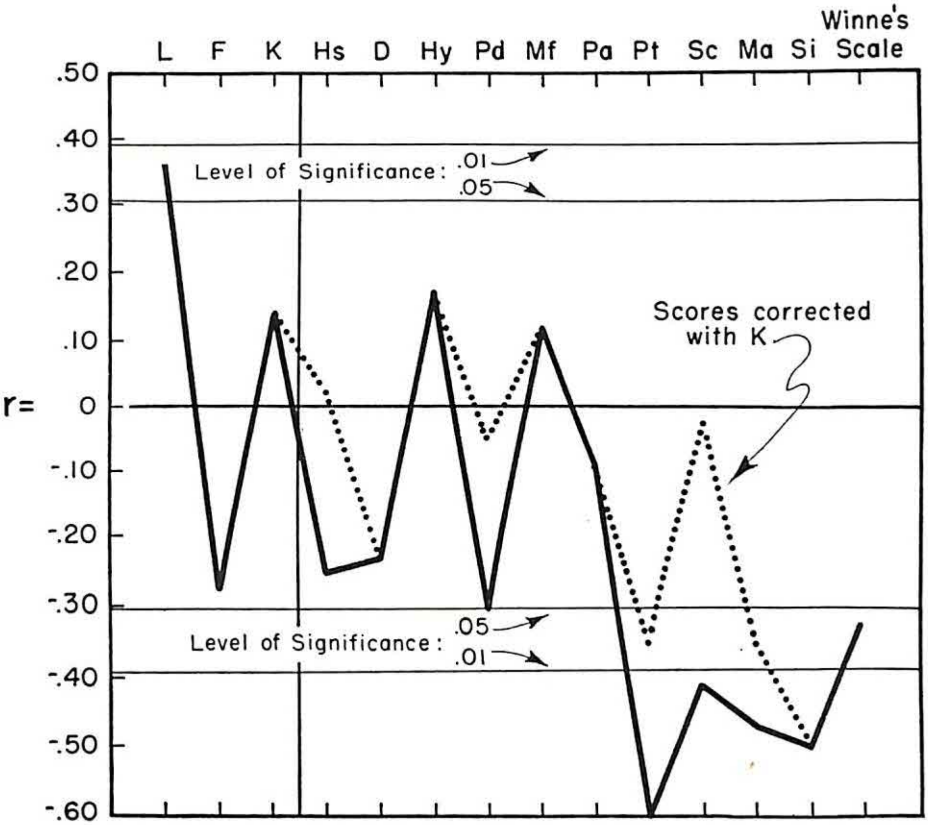

FIGURE 3
FREQUENCY OF BURPING OF INFANTS AS RELATED TO MOTHERS' SCORES ON CONVENTIONAL *MMPI* SCALES

by the *MMPI*, it is almost necessarily related to a number of the others since the *MMPI* scales are so highly correlated (13, 15).

As noted earlier, a number of investigators have suggested a positive relationship between anxiety and emotional disturbances of the mother during pregnancy and the later gastro-intestinal reactions of the child. While measures of the maternal variables were obtained not during the pregnancy but when the infants averaged three months of age, the test-retest reliabilities for these variables are of a sufficiently high magnitude (2) that one would expect that the mothers had essentially the same relative standing during pregnancy. Certainly there would seem to be no reason to expect a negative correlation between measures of these variables obtained during pregnancy and those obtained some three to six months later, a relationship which it

Bureau Ednl. Psy. Research
COLLEGE
Dated
Acc. No

would be necessary to assume if the present findings were to be explained by this hypothesis.

E. Summary

The hypotheses investigated in the present study are that gastro-intestinal disturbances of the infant are positively related to the mother's level of anxiety and to the extent of her emotional disturbance. Two reactions were used as indices of gastro-intestinal disturbance, incidence of regurgitation and frequency of burping. The first was chosen as an index because of its use by Ribble (8), Montague (6), Sontag (9, 10) and Escalona (1), and the second because of the logical consequences of Ribble's report that emotionally disturbed mothers are unable to adequately "mother" their children and that inadequately mothered children swallow more air (8). Extent of the mother's anxiety and extent of her emotional disturbance were measured by the *MMPI*.

The hypotheses were not supported by the results of the present study. Moreover, the statistically significant relationships were in a direction opposite to that expected on the basis of the earlier reports. In the absence of statistical measures and tests of significance in the earlier studies, they would appear to provide no basis for qualifying the conclusions of the present study.

References

1. Escalona, S. K. Feeding disturbances in children. *Amer. J. Orthopsychiat.*, 1945, **15**, 76-80.
2. Hathaway, S. R., & McKinley, J. C. The Minnesota Multiphasic Personality Inventory. (Rev. ed.) Minneapolis: Univ. Minnesota Press, 1943.
3. Hopper, H. E., & Pinneau, S. R. Regurgitation in infants as related to psychological variables. (Paper presented at 1954 meeting of the Western Psychological Association, Long Beach, California.)
4. ———. Frequency of regurgitation in infancy as related to the amount of stimulation received from the mother. *Child Devel.*, 1957, **28**, 229-235.
5. Kristal, J. T., & Pinneau, S. R. The relation of palmar sweat to personality variables. (Paper read at Western Psychological Association, Long Beach, May, 1954.)
6. Montague, M. F. A. Constitutional and prenatal factors in infant and child health. *In* M. J. E. Senn (*Ed.*), *Symposium on Healthy Personality*. New York: Macy Foundation, 1950. (Pp. 148-175.)
7. Pinneau, S. R. A critique on the articles by Margaret Ribble. *Child Devel.*, 1950, 203-228.
8. Ribble, M. A. Infantile experience in relation to personality development. *In* J. McV. Hunt (*Ed.*), *Personality and the Behavior Disorders*, Vol. II. New York: Ronald Press, 1944. (Pp. 621-653.)
9. Sontag, L. W. The significance of fetal environmental differences. *Amer. J. Obstet. & Gynecol.*, 1941, **42**, 996-1003.
10. ———. Differences in modifiability of fetal behavior and physiology. *Psychosom. Med.*, 1944, **6**, 151-154.

11. SULLIVAN, H. S. The Interpersonal Theory of Psychiatry. New York: Norton, 1953.
12. TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. Exp. Psychol.*, 1952, **44**, 61-64.
13. TYLER, F. T. A factorial analysis of fifteen *MMPI* scales. *J. Consult. Psychol.*, 1951, **15**, 451-456.
14. WELSH, G. S. An anxiety index and an internalization ratio for the *MMPI*. *J. Consult. Psychol.*, 1952, **16**, 65-72.
15. WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. J. The internal structure of the *MMPI*. *J. Consult. Psychol.*, 1951, **15**, 134-141.
16. WINNE, J. F. A scale of neuroticism: An adaptation of the Minnesota multiphasic personality inventory. *J. Clin. Psychol.*, 1951, **7**, 117-122.

Institute of Child Welfare
University of California
Berkeley 4, California

The Journal of Genetic Psychology, 1958, **93**, 15-32.

SEX-APPROPRIATE BEHAVIOR IN ADOLESCENCE*

St. Christopher's Hospital for Children, Philadelphia, Pennsylvania

G. H. Zuk

A. Introduction

The earlier physical maturation of girls is now an accepted empirical fact. Not only does the onset of puberty tend to occur earlier in girls, but maturation proceeds at a quicker rate. This study makes an inquiry into whether the differences in onset and rate of physical maturation have correlates at the behavioral level, especially in that area which has to do with the development of sex-appropriateness of behavior. Does it also hold true that significant differences exist between the sexes in the acquisition of sex-appropriate behavior?

Kuhlen (4) has pointed out that ". . . the direction of the trends during the adolescent years (e.g., whether girls are becoming more masculine or feminine) has not been unequivocally demonstrated" (p. 216). Tryon (5) found that:

> During the period between ages 12 and 15, values for girls have undergone revolutionary changes; values for boys have undergone relatively minor changes, mainly in terms of slightly shifted emphasis. . . . In studying the sex differences revealed by these data one is impressed with the lack of steadfastness to values revealed by the girls as compared with the boys over this relatively short period of time. . . . There is much greater demand upon these girls as compared to the boys for flexibility, for capacity to readjust their ideals, for ability to reorient themselves to new goals (pp. 77-79).

Tryon asks the question: "What is the genesis of the vacillation that we have observed in the personality of girls? Does such vacillation reflect some general conative ambivalence that is more characteristic of women than men? Is women's status and function in our culture more confused, more ambiguous than man's?" (p. 79). Benedek (1), psychoanalytically oriented, attempts to explain the dynamics of this conative ambivalence in the following way: ". . . whereas the boy is still free to project his life into the future (during adolescence), meanwhile modeling himself after various ideals, the girl is already preoccupied with the problem 'Who will

---

*Received in the Editorial Office on April 22, 1957.

love me? Am I lovable as I am? Can I and must I improve myself and who will love me then?' " (p. 20).

The objectives of this study were threefold: (*a*) to determine the "direction of trends" in sex-appropriate behavior during the adolescent years from 15 through 17, (*b*) to determine some of the general characteristics of sex-appropriate behavior, and (*c*) to determine the strength of the relationships between sex-appropriate behavior and possible psychobiological components. The study derived its basic measure of sex-appropriateness from an extensive list of everyday behaviors in which adolescents participated.

B. Design

The subjects of the study were 51 girls and 48 boys of the Adolescent Growth Study.[1] It was possible to obtain retests on these 99 subjects over a period of three years (1935-1938) on the U. C. Activity Record, an inventory of approximately 200 everyday behaviors. The Activity Record was composed of four sections, (*a*) Things I did last week, (*b*) Where I went last week, (*c*) What I read last week, and (*d*) Radio programs I listened to. Each of these sections listed approximately 40 to 60 individual behaviors. Instructions to the subjects filling out the Inventory were as follows:

> The pages in this booklet contain a number of lists of activities; things that students like to do in their spare time. Will you please indicate the activities you prefer by marking the items according to the directions for each set of records. There is no time limit, work carefully.

The 99 subjects, part of a much larger group given the Inventory, filled it out at about 6-month intervals during the years from 1935-1938. In 1935 the subjects were of an average age of about 14½, about 15½ in 1936, about 16½ in 1937, and about 17½ in 1938.

1. *Scoring Procedure*

Some statistical work that had already been done on the Activity Records collected suggested a procedure whereby the sex-appropriateness of behavior from 1935-38 could be measured. Percentages of each sex (based on N's for males of 151 in 1936, 166 in 1937, and 118 in 1938; N's for females of 170 in 1936, 179 in 1937, and 118 in 1938) marking a given behavior listed in Parts *A* and *B* of the Activity Record had been compared by means of the critical ratio. Significant sex differences had been obtained with respect to many of the behaviors. For example, more girls than boys took care of their rooms and did arts and crafts work during the years from 1935-38, whereas more boys than girls worked on bicycles and boats.

---

[1] Thanks are due Dr. Harold E. Jones, Director of The Institute of Child Welfare, Berkeley, California, for an opportunity to explore the data on these subjects.

An empirical criterion was settled upon for determining whether a particular behavior was more sex-appropriate for boys or girls. For example, in order for a particular behavior to be considered relatively more sex-appropriate for boys, it was necessary that that behavior be participated in by a greater number of boys than girls in each of the many times the Activity Record was administered from 1935-38. Such a behavior was considered male sex-appropriate even though a sizable percentage of girls might also have participated in the behavior. Thirty-four behaviors on the Activity Record were found to be checked consistently (i.e., from 1935 through 1938) by a greater number of boys; 32 items were found to be checked consistently by a greater number of girls. Thus, 66 behaviors out of about 200 were found to be relatively more sex-appropriate to one sex than the other. The remainder of the behaviors were found to lack the necessary consistency over the three-year age-span and were therefore discarded.

Activity Records given in the Fall of 1936, 1937, and 1938 were individually scored. It was arbitrarily decided that each individual would receive a score of one point if he (or she) had participated in a behavior which was relatively more appropriate to his (or her) sex. It was also decided that an individual would receive a score of one point if he (or she) did *not* participate in (that is, check) a behavior which was relatively less appropriate to his (or her) sex. Thus an individual's final score was the sum of the sex-appropriate behaviors checked *plus* the sum of sex-inappropriate behaviors left unchecked. In those instances where individuals did not participate in sex-appropriate behaviors or participated in behaviors relatively more appropriate to the opposite sex, it was decided that the best procedure would be to give the individual no score whatsoever, or, in other words, a score of 0. It seemed best not to penalize sex-inappropriate behavior too heavily, since it was possible for a large percentage of individuals of one sex to participate in a behavior which was consistently participated in by a larger number of the opposite sex.

Behaviors on which sex differences were found on Parts *A* and *B* to be consistently significant at about the .01 level (i.e., a *CR* greater than 3.00) were considered more strongly indicative of the existence of a sex bias and thus were accorded a special weighting: an individual checking such a behavior was given an additional score of one point. Six of the 34 male sex-appropriate behaviors were found to satisfy such a criterion on Parts *A* and *B* of the Activity Record; males checking such behaviors were given an additional score of one point. Five of the 32 female sex-appropriate behaviors were found to satisfy the criterion on Parts *A* and *B*; females checking such behaviors were given additional scorings of one point. Consequently, males and females *not* checking these particular sex-appropriate behaviors suffered a loss of two points rather than one.

Since critical ratios were not determined for the sex differences in checking the behaviors of Parts *C* and *D* of the Activity Record, it was decided that for these sections a certain percentage difference (corresponding to a *CR* probably as great or greater than 3.00) between boys and girls checking would satisfy the criterion by which some of the behaviors would be specially weighted; the percentage difference selected was 20 per cent. The difference, by inspection of comparable differences upon which *CR*'s of behaviors from Parts *A* and *B* were based, was considered one which would usually occur less than one time in 100 by chance. Six of the 34 male sex-appropriate behaviors were found to satisfy the criterion (i.e., a sex-difference as great or greater than 20 per cent) on Parts *C* and *D*; males checking such behaviors were given additional scorings of one point. Three of the 32 female sex-appropriate behaviors were found to satisfy the criterion on Parts *C* and *D*; females checking such behaviors were given additional scorings of one point. All in all, 12 of the 34 male sex-appropriate behaviors and eight of the 32 female sex-appropriate behaviors were specially weighted.

Each of the 99 subjects received three scores for sex-appropriateness of behavior based on the several parts of the Activity Record given in the Fall of 1936, 1937, and 1938. In addition, the possibility of differences in the Sections (*A, B, C, D*) was considered. Scores based on the behaviors selected from Parts *A* and *B* were kept separate from scores based on the behaviors selected from Parts *C* and *D*. It was felt that the behaviors listed in Parts *C* and *D* of the Activity Record ("What I read last week" and "Programs I listened to") were in a sense "more passive" than the behaviors listed in Parts *A* and *B* ("Things I did last week" and "Where I went last week"). In this analysis, *1935* scores were compared with 1937 and 1938.

It may need pointing out that the mean scores of males and females were not directly comparable. The maximum score a boy could make was 79; the maximum score for a girl was 75. Despite this, the rating procedure plus the special types of behaviors listed on the Activity Record made it easier for girls to get higher scores on the average than boys.

2. *Psychobiological Correlates of Sex-Appropriate Behavior*

In order to discover some possible determinants (or influencers) of sex-appropriate behavior in males and females, several social, physical, temperament, and intellective variables were selected out of the bulk of data of the Adolescent Growth Study. Each subject was rated "high" or "low" on the following list of variables according as his or her score fell above or below the median of the particular sex group:

1. Socioeconomic status: based on ratings of home appearance in 1935.
2. Strength: based on the average static and dynamometric hand strength in 1936 and 1938.
3. Need for aggression, social contact, and achievement: combined judges' ratings made in 1940 based on observations of behavior during adolescence.
4. Masculinity-femininity of build: based on ratings of body build by three experienced observers in 1937 and 1938.
5. *IQ*: based on the score derived from the Stanford-Binet administered in 1937.
6. Interest in school and achievement in school: based on ratings of teachers in 1937 and 1938.
7. Endomorphy, mesomorphy, and ectomorphy: based on ratings of experienced observer from photographs taken in 1937.
8. Early-late maturation: based on ratio of skeletal to chronological age during adolescence. Ratios greater than 1 were generally considered to indicate early maturation.
9. Masculinity-femininity of interest: based on scores of the Strong Vocational Interest Test administered in 1937.
10. Expressiveness, sociability, exuberance, irritability, and social adjustment: based on ratings of at least two independent judges in 1938.
11. Direct ego damage, rejection of aspiration and escape from reality: psychoanalytically oriented variables based on scores derived from the U. C. Inventory (developed by Caroline McCann Tryon). Direct ego damage is defined as ". . . admits various characteristics about himself and his situation, the recognition of which would be painful to most people." Rejection of aspiration is defined as ". . . says he is not and declares he is glad that he is not attractive . . . not liked by peers, not coöperative with parents and teachers and so on". Escape from reality is defined as ". . . admits daydreaming about being great or famous." Individuals were rated "high" and "low" according to the amount of year-to-year fluctuation in these variables—i.e., to the consistency of level—during adolescence rather than according to absolute amount during any one year.

C. RESULTS

1. *The Direction of Trends in Sex-Appropriate Behavior*

Table 1a presents means, standard deviations, and differences between means of scores on sex-appropriateness of behavior for males and females taking the Activity Record in 1936, 1937, and 1938. Also included in this table are means, standard deviations, and differences between means of scores based on Parts *A* and *B* of the Record given in *1935*, 1937, and 1938; that is, scores excluding "reading and listening" behaviors. The procedure of taking the total score from all parts of the Record plus the part-scores from

TABLE 1a

MEANS, STANDARD DEVIATIONS, AND DIFFERENCES BETWEEN MEANS OF RAW SCORES MADE BY BOYS AND GIRLS ON A MEASURE OF SEX-APPROPRIATENESS OF BEHAVIOR

Scores based on Complete Activity Record

Groups	1936	1937	1938	diff. (1936, 1937)	diff. (1937, 1938)	diff. (1936, 1938)
Males (= 48)						
Mean	40.79	40.00	39.98	.79	.02	.81
SD	5.43	4.80	5.55	—	—	—
range of scores	29–54	30–51	29–52	—	—	—
Females (N = 51)						
Mean	52.98	53.00	54.55	.02	1.55**	1.57*
SD	4.02	4.08	5.47	—	—	—
range of scores	43–61	43–61	43–63	—	—	—

Scores based on Parts *A* and *B* of Activity Record

	1935	1937	1938	diff. (1935, 1937)	diff. (1937, 1938)	diff. (1935, 1938)
Males (N = 43 in 1935, 45 in 1937 and 1938)						
Mean	24.43	23.18	24.00	1.25*	.82	.43
SD	3.74	3.33	3.61	—	—	—
range of scores	15–33	17–29	17–30	—	—	—
Females (N = 43 in 1935, 51 in 1937 and 1938)						
Mean	28.38	31.11	32.15	2.73**	1.04*	2.77**
SD	3.74	2.65	3.38	—	—	—
range of scores	20–34	26–37	25–38	—	—	—

*Significant at .05 level.
**Significant at .01 level.

Sections *A* and *B* was followed consistently in the study for purposes of general comparison. Part-scores from Sections *A* and *B* were selected rather than part-scores from *C* and *D* because the behaviors listed on these sections were initially considered—although the analysis of data provided little support for this--more sensitive discriminators of sex-appropriate behavior.

Mean scores in the male group declined slightly from 1936 to 1937 with little or no change from 1937-1938. Indeed, year-to-year differences in mean scores based on Parts *A* and *B* of the Activity Record reflect a significant *decline* in sex-appropriate behavior (*P* less than .05) from 1935 to 1937. In other words, there seemed to be an actual downward trend in male scores for sex-appropriateness of behavior from ages 14 to 16, especially in the area of "less passive" behaviors such as listed on Parts *A* and *B* of the Activity Record.

The female levels of sex-appropriateness of behavior from 1936 to 1938 showed rather a reverse trend from that of the males. There was almost no increase from 1936 to 1937, but a statistically significant increase from 1937 to 1938. The trend in mean scores based on Parts *A* and *B* was similar to that based on the total inventory. Thus, sex-appropriate behavior in the females seemingly increased significantly from the 16th to the 17th year.

In Table 1b, absolute changes (i.e., disregarding direction) in scores from one year to the next in males were compared with absolute changes in scores in females. A *t*-test of the difference in average absolute changes made by males and females from 1936 to 1937 was significant at the .01 level. In other words, male scores tended to fluctuate in *either direction* during this particular period to a greater degree than female scores. *t*-tests comparing the absolute changes from 1937 to 1938 and from 1936 to 1938 were not significant, but in these cases also the greater absolute change was made by the male group. The single reversal of this trend occurred in scores based on Parts *A* and *B* of the Record. The average absolute change in male scores from 1935 to 1937 was lower (3.16) than the comparable average change in female scores (4.19). This reversal may possibly be partially attributed to the special characteristics of the behaviors listed on Parts *A* and *B* and partially to the one-year different span on which the changes were assessed (1935 to 1937).

2. *Year-to-Year Stability of Sex-Appropriate Behavior*

Listed in Table 2 are product-moment correlations of raw scores derived from the Activity Record administered in 1935, 1937, and 1938, both from the complete Record and from Parts *A* and *B* only. Correlations for

J-315
3.9.70

TABLE 1b

AVERAGE YEARLY CHANGE (DISREGARDNG DIRECTION) IN RAW SCORES OF SEX-APPROPRIATENESS OF BEHAVIOR FOR MALES COMPARED WITH FEMALES

	Scores based on Complete Activity Record								
	1936-1937			1937-1938			1936-1938		
Statistic	Males	Females	Diff.	Males	Females	Diff.	Males	Females	Diff.
Mean	5.08	3.11	1.97**	4.31	3.49	.82	4.75	4.50	.25
SD	3.83	2.68	——	3.18	2.61	——	3.83	3.33	——
range of scores	0–19	0–10	——	0–12	0–9	——	0–16	0–12	——
	Scores based on Parts *A* and *B* of Activity Record								
	1935-1937			1937-1938			1935-1938		
	Males	Females	Diff.	Males	Females	Diff.	Males	Females	Diff.
Mean	3.16	4.19	1.03	2.54	2.10	.44	——	——	——
SD	2.20	2.81	——	2.28	2.13	——	——	——	——
range of scores	0–9	0–14	——	0–8	0–8	——	——	——	——

**Significant at .01 level.

TABLE 2

PRODUCT-MOMENT CORRELATIONS OF SCORES ON SEX-APPROPRIATENESS OF BEHAVIOR BETWEEN YEARLY ADMINISTRATIONS OF THE ACTIVITY RECORD; ALSO, PART-WHOLE CORRELATIONS

	Scores based on Complete Activity Record				Scores based on Parts *A* and *B* of Activity Record			
	1937		1938		1937		1938	
Year	Males (48)	Females (51)	Males (48)	Females (51)	Males (48)	Females (51)	Males (48)	Females (51)
1935	—	—	—	—	.42**	.22	.23	.30*
1936	.20	.47**	.39**	.38**	—	—	—	—
1937	—	—	.46**	.67**	—	—	.46**	.49**
Groups	Part (A,B) to whole (A,B,C,D) correlations of scores on Activity Record							
Males		—.25		.02	—		—	
Females		—.01		.36	—		—	

*Significant at .05 level.
**Significant at .01 level.

males (complete Record) ranged from .20 to .46. These were not particularly high, as one might expect of the stability of sex-appropriate behavior during adolescence, but they do indicate an increase in the trend toward stability from 15 to 17 years.

Correlations for females ranged from .38 to .67, generally higher than those for males. Correlations indicate also an increase in stability of sex-appropriate behavior with age. Correlations based on Parts *A* and *B* of the Record showed the same trends in stability of scores; in these, however, male behavior with respect to sex-appropriateness showed a slightly greater year-to-year stability. Correlations for males ranged from .23 to .46, for females from .23 to .49.

Generally speaking, the evidence suggests low but increasing stability of sex-appropriate behavior in boys and girls from 15 to 17 years. That greater stability of this behavior may be somewhat masked in the present data is indicated in an analysis of scores derived from different parts of the Activity Record. Part-whole correlations were determined between the scores of the complete Record and the scores obtained from Parts *A* and *B*. Essentially, this procedure compared the scores obtained on Parts *A* and *B* with those obtained on Parts *C* and *D*. Parts *A* and *B* contained what might be called "performance of tasks" (relatively more active) behavior, whereas Parts *C* and *D* contained what might be called "reading and listening" (relatively more passive) behaviors. Part-whole correlations indicated that there was *little or no relationship* between the scores of *A* and *B* and those of *C* and *D*. Table 2 shows that there was even a slight tendency in males in 1937 for high scores based on Parts *A* and *B* to be associated with low scores based on Parts *C* and *D* (—.25), although the negative correlation disappeared in 1938, giving way to an almost zero relationship (.02). In females in 1937 there was a zero relationship between scores based on the different parts of the Record, but in 1938 the correlation became definitely positive (.36), indicating some degree of consistency between scores based on the separate parts. Thus only in the case of females in 1938 did it appear that sex-appropriate behavior in different areas of behavior became fairly consistent. But this was not true of the sex-appropriate behavior of males in that year, or of the sex-appropriate behavior in both sexes in 1937.

The evidence suggests that the sex-appropriateness of behavior during adolescence varies considerably from one area of behavior to the next. To use an example, a boy whose behavior may be strongly sex-appropriate in the area of athletics may exhibit weak sex-appropriate behavior in the area of reading interests or hobbies. There is evidence that the same "unevenness" of sex-appropriate behavior exists in girls, but to a lesser degree. To continue with the example, a boy poor in athletics (i.e., not engaging much in such games) might "compensate" by reading heavily about sports and other activities demanding physical skills.

3. *Year-to-Year Stability of Behaviors on Which Sex Differences Were Observed*

The 34 behaviors marked by a greater percentage of males, and the 32 behaviors marked by a greater percentage of females were ranked separately according to relative popularity in 1936, 1937, and 1938. The 18 male and 20 female sex-appropriate behaviors obtained from Parts *A* and *B* of the Record were ranked independently of each other and independently of the 16 male and 12 female sex-appropriate behaviors listed in Parts *C* and *D* of the Record. *N*'s on which these rankings were made included the whole upper and lower grade classes to whom the Record was given in 1936, 1937, and 1938.

The rank-order correlations listed in Table 3 ranged from .51 to .99, suggesting that the popularity (or "frequency of mention") of behaviors which differentiated the sexes tended to be highly stable from year to year during this period of adolescence. Correlations showing the consistent year-to-year popularity of *male* sex-appropriate behaviors with *females* (and vice versa) are also rather high. It appears, in other words, that male sex-appropriate behaviors which are most popular with the *boys* tend to be also relatively most popular with the *girls*. Conversely, female sex-appropriate behaviors that are most popular with *girls* tend also to be most popular with *boys*. In addition, there appears to be little or no difference in the magnitude of correlations based on the behaviors obtained from Parts *A* and *B* of the Record and those obtained from Parts *C* and *D*.

The lower section of Table 3 suggests that the relationship between the sexes with regard to the relative popularity of the sex-differentiating behaviors *increases with age*. Male sex-appropriate behaviors (obtained from Parts *A* and *B* of the Record) were ranked according to popularity with both boys and girls in 1936, 1937, and 1938. These rankings were then correlated and the correlations compared for each of the years from 1936 through 1938. The correlations increased with age, from .73 in 1936 to .91 in 1938. The same trend occurred when female sex-appropriate behaviors (obtained from Parts *A* and *B*) were ranked according to popularity with both boys and girls in the successive years and these rankings were correlated; the correlations increased from .64 in 1936 to .79 in 1938. Similar correlational trends also occurred when rankings of boys and girls were compared with

TABLE 3
RANK-ORDER CORRELATIONS ESTIMATING THE CONSISTENCY OF POPULARITY OF MALE AND FEMALE SEX-APPROPRIATE BEHAVIORS FROM 1936 THROUGH 1938

	Behavior on Parts *A* and *B*						Behavior on Parts *C* and *D*					
	Male sex-appropriate			Female sex-appropriate			Male sex-appropriate			Female sex-appropriate		
Yearly comparisons	M	F	M with F	M	F	M with F	M	F	M with F	M	F	M with F
1936-1937	.77	.96	—	.99	.97	—	.91	.93	—	.92	.94	—
1937-1938	.88	.97	—	.97	.97	—	.96	.95	—	.91	.99	—
1936-1938	.69	.93	—	.96	.98	—	.88	.92	—	.78	.92	—
1936	—	—	.73	—	—	.64	—	—	.83	—	—	.51
1937	—	—	.84	—	—	.68	—	—	.85	—	—	.73
1938	—	—	.91	—	—	.79	—	—	.84	—	—	.81

Symbols: *M*, male.
F, female.
M with *F*, male and female rankings compared and correlated.

Note: Frequencies on which ranks were based: (*a*) boys (151, 166, and 108 in 1936, 1937, and 1938, respectively); (*b*) girls (170, 179, and 118 in 1936, 1937, and 1938, respectively.)

respect to those male and female sex-appropriate behaviors listed on Parts *C* and *D* of the Record.

4. *Physical, Intellective, Social, and Temperament Correlates of Sex-Appropriate Behavior*

The next step in the study involved the association of the measures of sex-appropriate behavior with selected variables which might seem logically to relate in some manner to that behavior. These variables were listed, together with vital information as to who rated them and when, in a previous section. Using the phi coefficient as the measure of association, tables of intercorrelations of the variables were developed. For the girls, intercorrelations of the variables ranged from —.52 to .76. The majority of the correlations, however, ranged in the interval from —.20 to .20. For the boys, intercorrelations ranged from —.77 to .82, but again these were only extreme cases. The great majority of correlations within and between "families" ranged between —.20 and .20.

Table 4 lists correlations (using phi coefficient as the measure of association) between the "best estimate" of level of sex-appropriateness of behavior and the variables. This "best estimate" was derived in the following manner: if two or more of the scores of an individual were above the median of his group, the individual's sex-appropriateness of behavior was considered to be "high"; if two or more of the scores were below the median, the individual was considered "low." In this way, the "highs" and "lows" of sex-appropriateness of behavior could be compared with "highs" and "lows" on the variables.

Table 4 also lists correlations between "best estimate" of sex-appropriate behavior based on scores obtained from Parts *A* and *B* of the Record and the variables. Listed in addition to these are correlations between scores in individual years (1936, 1937, and 1938) and the variables. For the boys, correlations between "best estimate" and the variables ranged from —.27 to .23. For the girls, correlations between "best estimate" and the variables ranged from —.32 to .20. Correlations between "best estimate" based on scores obtained from Parts *A* and *B* were no stronger for either boys or girls than comparable correlations based on scores obtained from the complete Record. In addition, correlations between scores obtained in individual years and the variables were no stronger than correlations between "best estimate" and the variables.

Of some interest are the *directions* of the correlations over the three-year period from 1936 to 1938. For both males and females, sex-appropriate-

TABLE 4

Correlations (Using the Phi Coefficient as the Measure of Association) Between "Best Estimate" Plus Yearly Measures of Sex-Appropriateness and Selected Social, Physical, Intellective, and Temperament Variables

Variables	"Best estimate" (based on scores of Complete Activity Record)		"Best estimate" (based on scores of Parts *A* and *B*)		Yearly estimates (based on scores of Complete Activity Record)					
					Males			Females		
	M	F	M	F	1938	1937	1936	1938	1937	1936
Socioeconomic Status	—.14	—.04	—.18	—.04	—.32*	—.18	—.09	—.12	—.20	—.04
Strength	.18	—.08	.27	—.12	.14	.55**	—.03	—.04	—.12	—.04
Need for Aggression	—.23	—.16	—.18	—.04	—.23	—.09	—.27	—.12	—.04	—.12
Need for Social Contact	.14	.08	—.18	.20	—.05	—.18	.09	.04	.12	.12
Need for Achievement	—.18	.00	—.18	—.12	—.18	.00	—.27	—.04	—.04	—.12
Sex-appropriateness of Build	—.18	.16	—.09	.28	—.23	—.18	—.09	.20	.12	.12
I.Q. (Stanford-Binet)	—.27	.20	—.27	.04	—.28	—.09	—.36*	—.04	.12	.12
Interest in School	—.18	.08	—.18	.04	—.14	—.18	—.27	—.04	.20	.21
Achievement in School	—.27	.00	—.18	.04	—.23	—.28	—.36*	—.04	.12	.20
Endomorphy	—.09	.00	.09	—.04	.01	.09	—.09	—.04	.04	—.04
Mesomorphy	—.14	—.12	.18	—.20	—.05	.27	.00	.04	—.12	—.12
Ectomorphy	.23	.00	.04	—.04	.13	—.14	.14	.04	—.04	.20
Early Maturation	.09	.08	.09	.04	—.01	.18	—.09	—.04	.20	—.04
Masculine Interests (Strong test)	.18	—	.27	—	.23	.00	.00	—	—	—
Expressiveness	—.18	—	—	—	.05	.00	.00	—	—	—
Sociability	.00	—	.09	—	.14	.18	—.09	—	—	—
Social Adjustment	—.18	—	.36*	—	—.18	.00	—.09	—	—	—
Exuberance	.09	—	—.18	—	.01	.27	.09	—	—	—
Irritability	.09	—	—.18	—	.14	.09	.09	—	—	—
Direct Ego Damage	—	.08	—	.20	—	—	—	.12	.12	—.04
Rejections of Aspirations	—	—.32*	—	—.20	—	—	—	—.36*	—.28	—.12
Escape from Reality	—	—.24	—	—.12	—	—	—	—.20	—.20	—.12

*Significant at .05 level.
**Significant at .01 level.

Correlations based on *N*'s of 50 females, 44 males.

ness of behavior seemed to be negatively associated with socioeconomic status, with the need for aggression and with mesomorphy, but positively associated with need for social contact and with early maturation. Because of the particular statistic employed in estimating degree of correlation, several of them were of zero order and therefore no decision could be made as to trends of association. However, correlations based on the scores from Parts *A* and *B* of the Record established a probable negative relationship (although slight) between sex-appropriate behavior and the need for achievement. The positive correlations previously established with the need for social contact and mesomorphy did not hold up in the comparable correlations based on Parts *A* and *B*. The differences between the correlations were probably due to the special characteristics of the "performance of tasks" behaviors of Parts *A* and *B* of the Record. Since the correlations between the parts of the Inventory were generally of low order, it was fully within expectations that the directions of correlations between the scores of the full Record and the scores of the part Record might vary. Of greater interest than the correlations which went in the same direction from both males and females were those which did not. Strength correlated positively with male scores for sex-appropriateness of behavior, but negatively with female scores; masculinity of build correlated *negatively* with male scores, femininity of build positively with female scores; *IQ* correlated negatively with male scores, positively with female scores; interest and achievement in school correlated negatively with male scores, positively with female scores. Correlations based on scores of Parts *A* and *B* of the Record showed that need for social contact was associated negatively with male scores, but positively with female scores. Correlations based on scores of Parts *A* and *B* of the Record showed that need for social contact was associated negatively with male scores, but positively with female scores; and that endomorphy, mesomorphy, and ectomorphy correlated positively with male scores, but negatively with female scores.

Male and female scores were correlated with 14 common variables. However, six separate variables were correlated only with male scores, with the following results: masculinity, as measured by the Strong Vocational Interest Test, correlated positively, and so did irritability and exuberance. Expressiveness and social adjustment correlated negatively, sociability correlated zero. Comparable correlations of these variables with the scores based on Parts *A* and *B* of the Record reversed some trends. In this series, social adjustment correlated positively but irritability negatively with scores of sex-appropriateness of behavior.

Three separate variables were correlated only with female scores, with the following results: year-to-year fluctuations in direct ego damage correlated positively, year-to-year fluctuations in rejection of aspirations and escape from reality correlated negatively with scores of sex-appropriateness of behavior. The comparable correlations of these variables with those scores based on Parts *A* and *B* of the Record showed no reversals in direction.

D. Discussion

1. *General Interpretations*

The findings tend to support the conclusion that differences exist between the sexes in rate of development of sex-appropriate behavior which parallel differences between the sexes in rate of physical maturation. Boys did not match girls in increase in sex-appropriateness of behavior during the period of adolescence studied; nor did they demonstrate in their behavior the year-to-year consistency or stability of girls.

Low correlations were found between measures of sex-appropriateness of behavior and measures of physiological, intellective, social, and temperament functioning. Studies carried out by other investigators (2, 3) have reported correlations between these categories of variables of a similar order. Alternative explanations of this finding may be considered: (*a*) that the variables actually are poorly correlated with sex-appropriateness of behavior or (*b*) that there is greater correlation but not during the adolescent period studied (i.e., between 15 and 17). Additional investigation is needed for a clarification of this issue.

2. *Relationship of Present Findings to Previous Interpretations*

Some discussion based on the observations and interpretations of other investigators may be in order. Kuhlen's earlier quotation stated: ". . . the direction of trends during the adolescent years (i.e., whether girls are becoming more masculine or feminine) has not been unequivocably demonstrated" (4, p. 216). Although the findings of this study cannot be said to have "unequivocably demonstrated" the fact that girls' behavior is becoming more feminine, the evidence of this study generally points in that direction. Contrariwise, the evidence of this study would suggest that boys' behavior during the adolesent period studied is *not* becoming more masculine.

The findings of this study agree with Benedek's observation reported earlier. It will be recalled that she said, ". . . whereas the boy is still free to project his life into the future (during adolescence), meanwhile modeling himself after various ideals, the girl is already preoccupied with the problem, 'Who will love me? Am I lovable as I am? Can I and must I improve myself, and who will love me then?' " (1, p. 20). If Benedek's observation is valid, then one might predict a greater directionality and stability in the sex-appropriateness of behavior of adolescent girls as compared with boys. Reasoning here would dictate that since the adolescent girl is more concerned than the adolescent boy with her "lovableness," she is more concerned also with the sex-appropriateness of her behavior.

Tryon, it was reported earlier in this paper, found that values for girls between the ages of 12 and 15 underwent a great change, whereas values for boys remained relatively unchanged. She concludes: "What is the genesis of the vacillation that we have observed in the personality of girls? Does such vacillation reflect some general conative ambivalence that is more characteristic of women than men? Is women's status and function in our culture more confused, more ambiguous than man's?" (5, p. 79). On the other hand, Tryon's empirical finding was that "There is much greater demand upon these girls (i.e., between ages 12 and 15) as compared to the boys for flexibility, for capacity to readjust their ideals, for ability to reorient themselves to new goals" (p. 79).

May not the alternative explanatory hypothesis from Tryon's be entertained as an interpretation of the results which she obtained? In other words, could not the changes in values in girls reflect—rather than a tendency to confusion as to sex rôle—a tendency to earlier acceptance and recognition of their adult sex rôle? The findings of the present study with special reference to the increase in sex-appropriateness of behavior in the adolescent girl group would lend support to this latter interpretation.

E. SUMMARY

The present study involved an analysis of everyday behaviors of 99 boys and girls of the Adolescent Growth Study for the purpose of investigating changes in sex-appropriateness of behavior during the adolescent years from 15-17. The results of the study may be summarized as follows:

1. Sex-appropriate behavior increased significantly in girls from 16 to 17 years. Boys showed no comparable increase during this period.

2. Sex-appropriate behavior tended to be more stable in girls from year to year than boys. For both sexes, however, such behavior was more stable during the 16th than the 15th year.

3. Behaviors which were sex-appropriate and more popular with boys tended also to be relatively more popular with girls, and vice versa.

4. Sex-appropriate behavior was shown to be related in reasonable directions but in low degree with social, physical, intellective, and temperament factors.

5. The sex-appropriateness of adolescent behavior was shown to vary widely from one area to another. It was not, for example, unusual for an adolescent boy or girl to exhibit at one and the same time high and low sex-appropriateness between "families" of sociosexual behavior.

REFERENCES

1. BENEDEK, T. Insight and Personality Adjustment. New York: Ronald Press, 1946.
2. CAMPBELL, E. H. The social-sex development of children. *Genet. Psychol. Monog.*, 1938, **21**, 461-552.
3. GILKINSON, H. Masculine temperament and secondary sex characteristics: A study of the relationship between psychological and physiological measures of masculinity. *Genet. Psychol. Monog.*, 1937, **19**, 1-104.
4. KUHLEN, R. G. The Psychology of Adolescent Development. New York: Harper, 1952.
5. TRYON, C. M. Evaluations of adolescent personality by adolescents. *Monog., Soc. Res. Child Devel.*, 1939, **4**, No. 4.

St. Christopher's Hospital for Children
Lawrence and Huntingdon Streets
Philadelphia, Pennsylvania

The Journal of Genetic Psychology, 1958, **93**, 33-36.

EFFECTS OF SATIATION ON FOOD PREFERENCES IN MONKEYS*[1]

Department of Psychology, Stanford University

J. M. WARREN

A. PROBLEM

Harlow *et al.* (1, 3) have demonstrated that monkeys' food preferences may be scaled reliably by the method of paired comparisons, that the order of preference is not affected by differential food deprivation (1), and that rankings of individual foods may be reproduced after a three-month interval (3).

The purpose of this experiment was to determine whether the ordinarily stable food preferences of monkeys could be modified by selective satiation.

B. METHOD

1. *Subjects*

Eighteen preadolescent rhesus monkeys were studied. Their previous laboratory experience consisted of 10 days' testing for lateral preference in the Wisconsin General Test Apparatus (as in 4).

2. *Apparatus*

The W. G. T. A. has been described previously (2, Figure 1). The foods were presented on a tray containing two depressions, 2½ in. long, ½ in. wide, and 1/16 in. deep, which permitted efficient visual discrimination.

3. *Procedure*

Five different foods were used: peanut, raisin, potato, celery, and egg mash pellets. Spanish peanuts and raisins are naturally about the same size; potato and celery cubes were cut to approximately the same size as the former foods. The egg mash pellets were about half the size of the other foods, so that two pieces were considered equivalent to the arbitrary standard unit.

On each trial, one kind of food was placed on one side of the tray and another food on the opposite side. Then the tray was advanced within reach of *S* and a single choice was permitted. Failure of response was recorded if *S* failed to take either food within 30 seconds.

---

*Received in the Editorial Office on April 29, 1957.

[1] This research was supported in part by Grant M-835 from the National Institute of Mental Health, National Institutes of Health, and was conducted at the University of Oregon.

The positions of the different kinds of food were randomized, and the possible combinations of five things taken two at a time appeared in an irregular order. Each comparison appeared once in every block of 10 presentations. Fifty presentations were given each day, five days per week (test period).

The monkeys were tested under two motivational conditions. In four of the five test periods, the animals were tested under standard deprivation conditions, i.e., 23 hours after the last feeding. During the satiation period, the *S*s which had been deprived for 23 hours were given unlimited access to a large quantity of peanuts for 15 minutes immediately before testing. Group *A* (six monkeys) were tested under standard conditions during Periods I and II, then the satiation procedure was introduced during the third test period (*S*), and then these animals were tested under standard conditions again during the post-satiation (*PS*) period.

Group *B*, which consisted of 12 animals, was tested under 23 hours deprivation for three weeks (Periods 1, II, and III) before the satiation period (*S*).

The monkeys were next tested in paired social competition for four weeks (5). Following this study of social dominance, all *S*s were again tested for food preferences for five days. This period, following the dominance tests, is identified as *D* for both groups.

C. Results

The percentage of trials on which each food was taken in preference to all other foods is plotted against successive test periods in Figure 1. Each point for Group *A* is based upon 600 and for Group *B* upon 1,200 trials. Those trials involving low preference foods on which no response occurred within 30 seconds were recorded as refusal of both foods, since it seemed desirable to study absolute as well as relative acceptability of the foods.

In both groups, raisin and peanut were the most highly preferred foods, and celery the least preferred. Potato and egg mash were intermediate in preference value, and acceptance of these foods varied somewhat in time and between groups.

During the prefeeding session immediately prior to testing during the satiation period, the monkeys took an average of 23 grams of peanuts—a quantity too large to be eaten immediately, so that almost all animals were

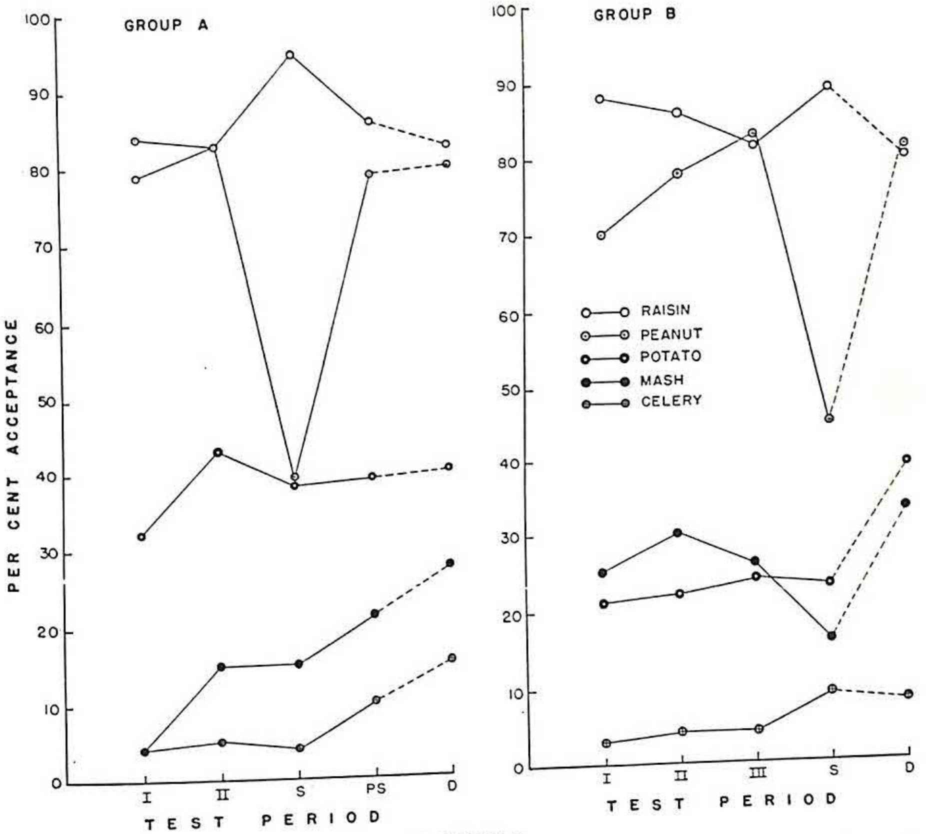

FIGURE 1

Per Cent Acceptance of Test Foods on Successive Test Periods by Groups *A* and *B*

tested with their cheek pouches bulging with the surplus. In spite of this apparent surfeit of peanuts, however, the effect of prefeeding was not such as to alter the order of preference markedly in either group. Acceptance of peanuts decreased by about 40 per cent and acceptance of raisins increased by about 10 per cent in both groups during the satiation periods, but in spite of the decreased acceptance of peanuts this food was still preferred over potato, egg mash, and celery. The data for Group *A* indicate that acceptance of peanut returned to approximately pre-satiation level in the post-satiation tests during the following week. No permanent effect of satiation is apparent, since both groups readily accepted peanuts during Period *D*.

D. Discussion

The results of this experiment indicate that feeding large quantities of a high preference food before determining differential preferences by monkeys transiently increases the acceptability of another highly preferred food, and decreases temporarily the acceptability of the food available before the test and of low and intermediate preference foods. However, the order of preference is essentially unchanged by this procedure, and acceptance levels for individual foods return to pre-satiation levels upon the discontinuation of the prefeeding procedure. These findings, together with the great similarity of the results obtained from two groups tested under somewhat different experimental conditions, confirm the previous reports (1, 3) emphasizing the consistency and reproducibility of food preferences in monkeys.

E. Summary

The relative acceptability of peanut, raisin, potato, egg mash pellets, and celery for 18 rhesus monkeys was determined by the method of paired comparisons. The following results were obtained:

1. With 23 hours' food deprivation, raisins and peanuts were taken much more frequently than the remaining foods; celery was the least preferred food.
2. Prefeeding with peanuts depressed the acceptance of peanuts by about 40 per cent, increased the acceptance of raisins by about 10 per cent, and generally depressed the acceptance of the less preferred foods.
3. Prefeeding with peanuts had no permanent effect on the acceptability of this food.

References

1. Fay, J. C., Miller, J. D., & Harlow, H. F. Incentive size, food deprivation, and food preferences. *J. Comp. & Physiol. Psychol.*, 1953, **46**, 13-15.
2. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51-65.
3. Harlow, H. F., & Meyer, D. R. Paired-comparison scales for monkey rewards. *J. Comp. & Physiol. Psychol.*, 1952, **45**, 73-79.
4. Warren, J. M. Handedness in the Rhesus monkey. *Science*, 1953, **118**, 622-623.
5. Warren, J. M., & Maroney, R. J. Competitive social interaction between monkeys. *J. Soc. Psychol.*, (in press).

Department of Psychology
Stanford University
Stanford, California

The Journal of Genetic Psychology, 1958, **93**, 37-51.

THE PERFORMANCE OF X-RAY IRRADIATED AND NON-IRRADIATED RHESUS MONKEYS BEFORE, DURING, AND FOLLOWING CHRONIC BARBITURATE SEDATION*[1]

Department of Psychology, University of South Dakota, and the Sacred Heart Hospital, Yankton, South Dakota

R. T. Davis, A. A. McDowell, M. A. Grodsky, and J. P. Steele

A. Problem

Six months before the present study commenced each *S* in one of the two groups of monkeys employed in this study had received a whole body dose of 400 *r* *X*-ray irradiation. Dramatic physiological changes occurred following this treatment, but as an earlier paper indicated (3) the performance of *S*s on a wide variety of laboratory tasks changed surprisingly little after irradiation, and very few differences were noted between the performance of the irradiated *S*s and the performance of the *S*s serving in a nonirradiated control group. The present study employed the same *S*s as the previous study, used nine of the 10 original tasks, and investigated the performance of these monkeys before, during, and after an 18-day period of daily barbiturate sedation. The purpose of this procedure was to contrast debilitation induced by chronic barbiturate sedation with debilitation following whole body irradiation and to determine the possibility of identifying chronic radiation effects through the synergic action of barbiturate sedation.

Furchtgott (4) suggested that the behavioral changes following whole body irradiation can be reduced to a general malaise. If this is the case, other debilitating agents might produce a similar syndrome. Barbiturates in surgical doses have been given to monkeys by Weiner and Harlow (10). These *E*'s trained their *S*s before sedation, administered sedation, then resumed training their *S*s as soon as they had recovered sufficiently from anesthesia to respond. Training was continued for six hours thereafter. These workers showed that the performance of rhesus monkeys on standard laboratory tasks was severely affected several hours following the *S*s' recovery from sedation.

The chronic dose used in the present study was decided upon in part be-

---

*Received in the Editorial Office on May 3, 1957.

[1] This investigation was supported in part by a research grant (M-530, C-2) from the National Institute of Mental Health.

cause of the inherent danger to *S*s of surgical doses and in part because of *E*'s desire to produce debilitation for a period of time corresponding to that of the greatest biological pathology following irradiation.

B. METHODS

1. *Subjects*

The *S*s employed in this study were 15 rhesus monkeys with nearly identical training during their previous 22 months in the South Dakota laboratory. Nine *S*s had received 400 *r* irradiation with *X*-rays six months before the present study began and the remaining six had not been irradiated, serving as a normal control group. Details of the previous taming, training, and irradiation procedures have been reported elsewhere (2, 3, 7, 8).

2. *Apparatus*

The nine procedures were carried out in either the *WGTA* or in the *S*s' living cage. The living arrangements and radiation apparatus and procedures have been described previously (3).

3. *General Procedure*

As reported earlier (3), each of 16 *S*s had received training on 10 widely different laboratory tasks during a period of a year. Then 10 *S*s were each given 400 *r* whole body irradiation with *X*-rays and six *S*s formed a nonirradiated control group. Following the radiation of the experimental group the six control and the nine surviving experimental *S*s were studied 80 days on the same tasks they had received prior to the period of radiation.

After completion of the postradiation period of study, *S*s were confronted with two object preference tasks which have been described elsewhere (1, 5). These tasks were dissimilar to any tasks of the original study and occupied a four-month period between the end of the postradiation training period and the initiation of the present study.

Nine of the original tasks were employed in the present study. These tasks were given in three batteries and one of these batteries of tasks was given each day of the training period. The training schedule consisted of a 24-day period of presedation training during which each task was given eight times in a predetermined random order to each *S*, an 18-day intrasedation period during which animals each received 0.5 gm. phenobarbital sodium crushed into 100 gm. banana once a day, and an 18-day postsedation period. The time of dosage preceded the daily period of training and observation by two hours and the time a *S* received his daily dosage of phenobarbital was randomized throughout the sedation period. Each battery of tasks was given to each *S* on six different occasions during the intra- and on six occasions during the postsedation period. Retention by *S*s over a period of four months for the original tasks was measured by comparing performance of *S*s during the last 32 calendar days of the 80-day postradiation period with the performance of *S*s during the presedation period. During this arbitrary part of the postradiation period each task was given to each *S* on eight occasions. The nine tasks common to both experiments included the original six manipulation tasks designated: (*a*) the bent-wire task with Life Saver reward, (*b*) the bent-wire task with poker chip reward, (*c*) the simple puzzle position tasks, (*d*) the six-unit mechanical puzzle task, (*e*) the patterned-string task, and (*f*) the wood-block chewing task.

Three of the original tasks required *S*s to make correct stimulus-object choices for rewards of food: (*g*) the four-trial object-quality discrimination learning task, (*h*) the five-sec. delayed response task, and (*i*) the reduced-cue discrimination problem task. Since the procedures of these tasks were common to the original radiation study and the present experiment, very little procedural detail will be included in this paper and only necessary description of trial distribution will be presented. Tasks will be grouped and analyzed according to the categories of manipulation tasks and object-discrimination tasks. The results were analyzed by employing the sign test (9) based on comparisons of raw scores of *S*s recorded during the last part of the postradiation period and during each of the three periods of the present study. Scores of each *S* in the irradiated experimental group ($N = 9$) were compared with his scores in subsequent periods of the experiment, and the performance of each *S* in the control group ($N = 6$) in two adjacent periods of these experiments was compared. Finally the performance of each *S* in the pooled experimental-control group ($N = 15$) in one stage of training was compared with each *S*'s performance in subsequent stages of training.

C. Specific Procedures and Results

1. *Manipulation Tasks*

a. The bent-wire Life Saver task. The 40 basic patterns described previously were employed. The construction of the patterns in the trial procedure was the same as that described in the radiation study (3). Five replications of these patterns were given, two replications in the pre-, one in the intra-, and one in the postsedation period of training. The order of pattern presentation was random and a complete replication of all 40 patterns required six days, five days to present 35 problems, and the sixth day to present the remaining five problems. These six occasions of training were spaced once every three days for 18 days. The *S*s' behavior occurring during the present study is contrasted with the performance of the same *S*s on one replication of 40 patterns recorded at the end of the postradiation period four months prior to this study.

The performance of *S*s on the bent-wire detour problems was compared at four periods: at the end of the radiation study, during the presedation period, while receiving chronic barbiturate sedation, and during the postsedation recovery period. Performance was measured as the mean time to remove the Life Saver reward rather than the number of Failures or Failures-after-Trying, because all *S*s had mastered the total range of problem difficulty and only three failures in 3,000 trials occurred in the present study. Subjects in neither the irradiated nor control groups forgot the principle of solving these problems between the last postradiation and the first presedation replication of 40 patterns.

Sedation did not have a very striking effect on performance although eight of the nine experimental *S*s took a longer time to remove lures from the easiest patterns (1-8) during the period of sedation. Generally the nonirradiated *S*s took less time to remove the lure during the sedation period than in the presedation period, and this difference was significant for Patterns 17 through 24. On the more difficult patterns the irradiated *S*s also responded more rapidly during the sedation than during the presedation period but in no instance was this difference significant.

b. The bent-wire poker chip task. The 12 bent-wire problems employed previously in the bent-wire poker chip task (3) were used in the present study. Six randomly preselected patterns were given each *S* during eight days of the 24 presedation training days, during six of the 18 days of the intrasedation training and during six days of the 18 days of postsedation training. The problems were given to each *S* in its living cage in the manner previously described and *S* received paper poker chips as rewards for threading the punched poker chips off the wire pattern. Three response categories were employed: (*a*) *responses failed*, (*b*) *responses attempted by manually manipulating the potential reward*, and (*c*) *responses attempted by chewing the poker chip*. The latter two categories are used in the place of the previously defined category of *failed-after-trying*.

Figure 1 shows the performance of the *S*s in the irradiated and nonirradiated groups during the three stages of the present study: pre-, intra-, and postsedation. The subfigures show, from left to right, (*a*) per cent attempted responses, manual, (*b*) per cent attempted responses, oral, and (*c*) per cent

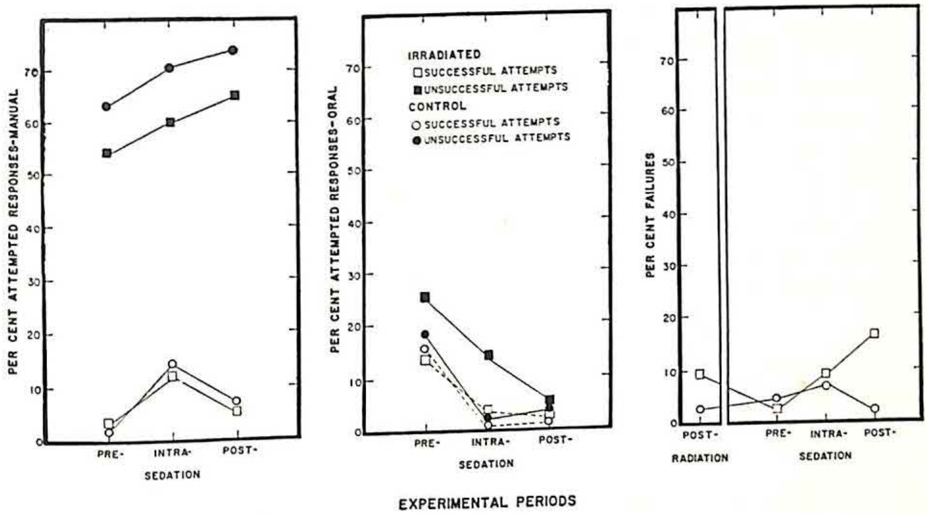

FIGURE 1

Performance of normal and *X*-ray irradiated monkeys on the bent-wire poker chip task before, during, and after a period of chronic sedation. Performance of animals in the experimental group is designated by squares and control group by circles. In the case of attempted responses, either manual or oral, shown in the left hand and center parts of the figure, the open circles and squares indicate successful responses and the filled circles and squares indicate unsuccessful responses.

failures to respond. The *S*s failures during the last period of postradiation training are also shown. The figure indicates a number of changes during the course of the sedation study. The performance of both groups of *S*s at the beginning of the present study is comparable to that at the end of the postradiation period. Throughout the three stages of the sedation study three different sources of change in performance are evident. The first type of change is primarily due to the introduction and removal of the sedation condition. This is exemplified by the significantly more frequent successful manual responses during the sedation period than during the presedation period. The postsedation fall in performance barely misses being significant. A significant rise occurs between the pre- and intrasedation scores of all 15 *S*s, but the change is not significant when the scores of the *S*s in either the experimental or control group is examined separately.

The second change is quite possibly a satiation effect. A general decline in oral manipulation (oral *FT* as well as oral success) is noted throughout the three stages of this experiment. This trend is significant when *E* compared the performance of *S*s during the pre- and intrasedation periods, and the difference is significant for the *S*s in the pooled group as well as the *S*s in the irradiated group alone.

The third change is clearly an interactive effect between previous irradiation and present sedation. Between the pre- aand postsedation periods, and developing through the intrasedation period, is a tendency for *S*s in the irradiated group to fail to respond. Only *S*s in the irradiated group have significantly fewer failures in the post- than in the presedation training periods, and this trend appears very similar to that noted in the performance of the irradiated *S*s immediately following irradiation (3).

c. The simple-puzzle position task. Four simple mechanical devices were used. One contained nine eye bolts inserted in holes in a 7 by 10 by ¾ in. gray-painted plywood board, the second included nine 2-in. bolts and eyes, the third nine hasps and hasp staples, and the fourth was a composite of the elements in the first three devices and consisted of three successive sets of units. A set was comprised of a hasp restrained by a hook which in turn was restrained by a screw eye. These puzzles were mounted either on the test tray of the *WGTA* or in the *S*s' living cage. Each of the four devices was presented to a *S* in a random order for a 2-min. period on every day of puzzle presentation. The distribution of training was the same as for the two bent wire tasks, and all four puzzles were given on one of three days constituting a training cycle during the 12 training cycles of the present study.

Numerous and complex changes took place in the performance of *S*s on the simple puzzles. These changes can be seen by examining Figures 2 and 3 which represent the performance of *S*s in both the experimental and control groups on each of the four puzzles. The performance of *S*s in the *WGTA* shown in Figure 2 is separated from the performance in the living cage shown in Figure 3 and the performance on each of the four puzzles is shown separately within each figure. To facilitate the reader's examination of the figure *E* has indicated each significant rise or fall in the figure legend.

An over-all appraisal of the figures indicates that *S*s in the control group never registered a significant rise or fall in any of the comparisons made. However, the performance of previously irradiated *S*s rose or fell significantly in seven of the 24 comparisons made between the performance of the irradiated *S*s on successive stages of this study. These changes were primarily increases in manipulation during the period of sedation and declines in manipulation after sedation was removed. The latter change was restricted to the performance of *S*s on puzzles that were presented in the living cage. Significant changes in performance were recorded on all four puzzles although more changes occurred in the performance of *S*s on Puzzle 4 than on the

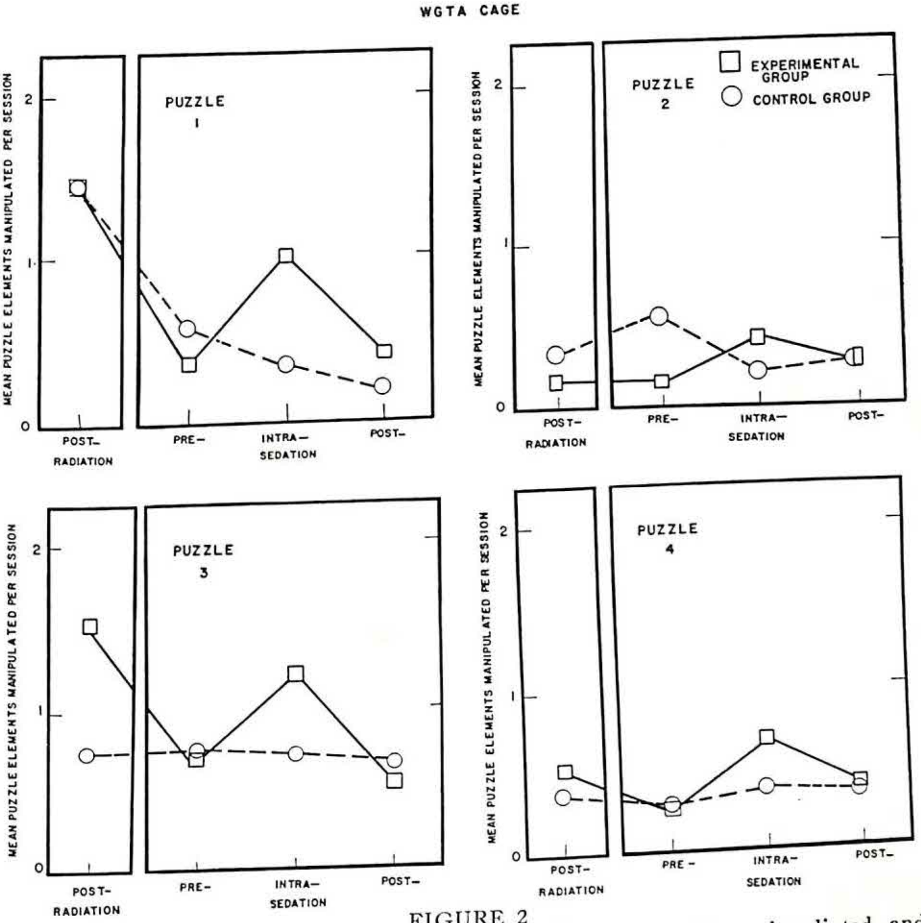

FIGURE 2

Mechanical puzzle performance in the *WGTA* of *S*s in *X*-ray irradiated and control groups at the end of the period following prior irradiation, and during the three periods of the sedation study. The squares connected by solid lines represent the performance of *S*s in the experimental group and the circles connected by the broken lines indicate the performance of *S*s in the control group. The increase in manipulation by the *S*s in experimental group between the pre- and intrasedation period performance is significant on Puzzle 1 and on all puzzles taken together.

other puzzles. Subjects in the previously irradiated group tended to manipulate puzzle elements less in the presedation period than in the last eight presentations of puzzles in the postradiation period, but this trend was significant in only one comparison; the irradiated *S*s manipulated the elements of Puzzle 4 presented in the living cage significantly less after the lapse of time between the postradiation and the beginning of the presedation period.

d. The six-unit mechanical puzzle task. The six-unit puzzle was used

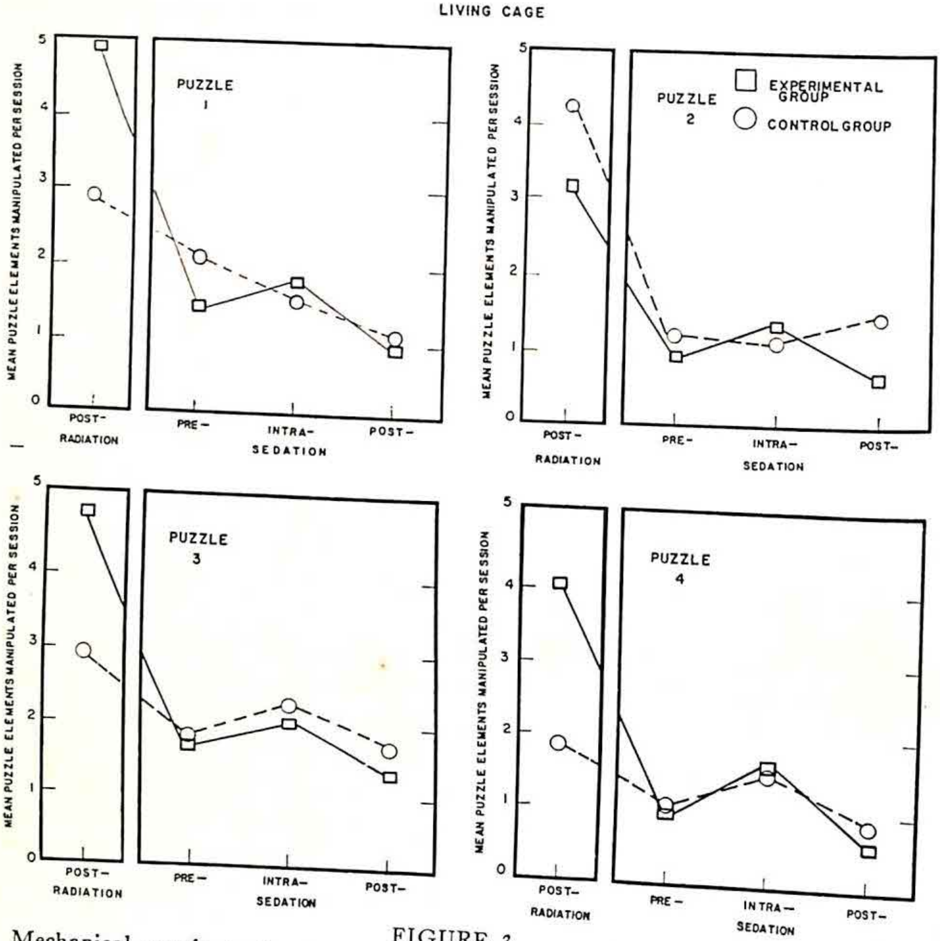

FIGURE 3

Mechanical puzzle performance in the living cage of *S*s in *X*-ray irradiated and control groups as a function of chronic sedation. The figure represents *S*s' performance at the end of the period following prior irradiation and during the three periods of the sedation study. The squares connected by the solid lines represent the performance of *S*s in the experimental group and the circles the performance of *S*s in the control group. *S*s in the experimental group made significantly fewer manipulations of Puzzle 4 in the presedation period than in either the postradiation period or in the intrasedation period and significantly fewer manipulations in the postsedation period than during the intrasedation period. The *S*s in this group also made significantly fewer manipulations of Puzzle 2 during the post- than during the intrasedation period. Significant declines in performance between the intra- and postsedation periods were indicated on Puzzles 2 and 4 for all 15 *S*s pooled together.

in the previous radiation study (3) and is very similar to a puzzle used originally in the Wisconsin laboratory. The puzzle was attached to the bars of the living cage door by means of a 1- by 6-in. wooden turn screw and was presented 15 minutes daily during the time the *S*'s cage mate was being tested elsewhere. Each *S* received the puzzle on eight occasions in the presedation period, six times in the intra-, and six times in the postsedation training periods.

Analysis of the data shows no significant changes in the performance of *S*s in either group or in the pooled experimental-control group as a function of introducing or halting sedation. Furthermore, there is a striking similarity between the performance level of *S*s in both groups during the last eight days of the postradiation period and during the presedation period.

e. The pattern-string problem task. The cross, box, angle cross, four crossing, double cross and pseudo-cross string pattern problems were each employed once daily in a random order on the 20 days that the first test battery was given. A raisin or quartered grape reward was impaled on the end of the correct string. The methods used have been described in detail elsewhere (3).

The performance of *S*s on the pattern-string task before, during, and after sedation showed no systematic changes and the level of performance was comparable to that occurring at the end of the postradiation period.

f. The wood-block chewing task. At the beginning of the presedation period a wood block composed of two 3 by 3 by 1 inch pieces of pine board bolted together with a 2.5-in. eye bolt was introduced to each *S*. The block was presented 15 min. each day that the third battery of tasks was given in the pre-, intra-, and postsedation periods. Details of this task have been presented elsewhere (3, 7). Figure 4 shows the performance of *S*s during the last eight of the 20 postradiation days that a block was presented, and the performance of *S*s after a new block was presented at the beginning of the present experiment. Inspection of the figure and legend shows that performance of *S*s in neither the experimental nor control group taken alone showed a significant change between the last eight postradiation days and the eight presedation days; but if *S*s in the experimental and control groups are pooled, the over-all group shows a significant increase in chewing during this period. An increase in chewing was previously noted when a new block was introduced (7) and the *S*s' significantly greater chewing at the beginning of the present experiment may be attributed to the introduction of a new block. During the intrasedation period, *S*s, particularly the control group *S*s, chew blocks more vigorously than in the presedation period. This difference is

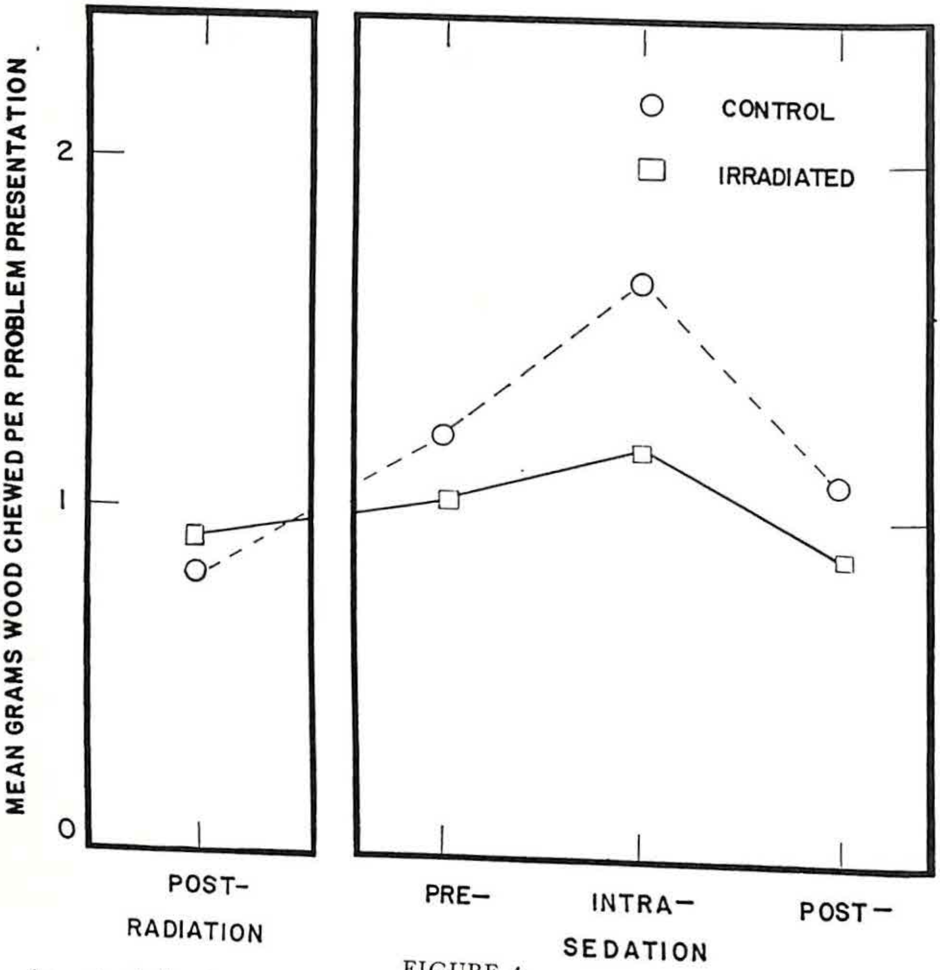

FIGURE 4

Amount of chewing on a wood block as a function of sedation of monkeys who had received prior high dose irradiation with *X*-rays and normal control *S*s. The *S*s in the control group chewed significantly less wood off the blocks during the postsedation period than during the intrasedation period and the pooled group of control and irradiated *S*s showed a significant increase in chewing between the postradiation and presedation period.

occasioned by more chewing of five out of six controls and six out of the nine experimental animals and is not significant. However, the direction of this trend is in the opposite direction of any trend in chewing performance previously noted and it appears that sedation interfered with the usual decline in the *S*s' chewing noted between first and second periods of time the same wood block is presented. The *S*s in the control group chew the wooden block significantly less during the postsedation period than during the intrasedation period and *Ss* in the pooled experimental-control group chew significantly less. This satiation trend is similar to that noted previously (7).

2. *Object Discrimination Tasks*

Three of the four object-discrimination tasks used in the original procedures were employed in the present experiment. The results of the present study were similar to those found in the radiation study and indicated that neither the introduction nor the removal of the sedation condition produced significant changes in the performance of *Ss* in either the experimental or control groups. The following trial schedules were followed in presenting the object discrimination tasks.

a. The four-trial object-quality discrimination learning task. Subjects were given six 4-trial object-quality discrimination problems a day on the eight days that the first battery of tasks was presented in the presedation periods. Each *S* received six problems a day for the six applicable intra- and six applicable posttraining days. Training was conducted in the *WGTA* and followed the training paradigm used by the Wisconsin workers (11). Analysis indicated no significant changes in performance of either or both groups between the final eight days' training postradiation and the beginning of the presedation period. Similarly no significant changes occurred in *Ss*' performance during the course of the present experiment.

b. The delayed-response task. Eight 5-sec. and eight 20-sec. delayed-response problems were given to each *S* during eight days spaced throughout the presedation training period and during six days of the intra- and six days of the postsedation period. The procedures differed from those previously given in that a pair of identical orange painted wooden blocks were utilized on every trial and the unfamiliar pairs of blocks used in the radiation study were omitted. Trial procedures were similar to those used by Harlow and his co-workers to present delayed-response problems in the *WGTA* (6). Performance of *Ss* on delayed-response problems was not significantly altered by the present procedures.

c. The reduced-cue discrimination problem task. This task has been described in considerable detail elsewhere (2, 3) and represents a compromise between delayed-response and object-quality discrimination problems. The problem was given in the *WGTA* and a two-food well stimulus tray was employed. A problem consisted of cue-trials in which two identical objects were placed over the food well. One of these objects was arbitrarily rewarded, and the other was not rewarded. The position of the reward was indicated by a third object placed upon the correct object. On the second trial Ss received only the two identical stimuli and to solve the problem displaced the object in the position of the previously correct object. The performance of the nonirradiated Ss in the control group fell significantly during the interval between the last eight times this problem was presented in the postradiation period and the beginning of the presedation period of the present experiment. This trend was not significant for Ss in the experimental group but was significant when the performance of the pooled groups was compared.

D. Discussion

The results can best be discussed in terms of the four types of changes in performance that were possible within the scope of this experiment: (*a*) changes of retention, (*b*) the effect on performance of chronic sedation, (*c*) changes in performance during or after sedation which were radiomemetic, (*d*) changes in performance of Ss which were radioantithetic.

1. *Retention*

Two changes in performance of Ss occurred after the interposed rest interval. One of these changes involved an increase in Ss' errors on the reduced-cue discrimination problem between the postradiation and presedation periods and possibly represents forgetting. The other change was manifest by Ss in the irradiated group as less puzzle manipulation in the presedation period than in the postradiation period on one of the puzzles of the simple puzzle position task. Since the amount of manipulation of Ss in the irradiated group is considerably greater than that of Ss of the nonirradiated group in the last part of the postradiation period, this change is possibly a return to normal by Ss in the irradiated group rather than forgetting or deterioration of a skill. The relatively few significant instances of forgetting over a four-month lapse of time is not surprising since Ss had been overtrained and the tasks employed did not involve the memory of particular discrimination but rather of concepts (i.e., object-quality discrimination learning sets), of motor skills (i.e., bent-wire detour problems) or non-learned skills (wood-block chewing task).

2. *Sedation Effects*

The effects of sedation *per se* were limited. The number of manually attempted responses made by Ss on the bent-wire detour problem with poker chip reinforcement increased during the period of sedation and fell thereafter; this change affected all Ss and did not appear to interact with the radiation variable. Sedation also appeared to interfere with expected changes in the satiation of chewing responses; but the remaining changes accompanying sedation seemed to involve an interaction between the effects of prior irradiation and sedation.

3. *Radiomemetic Sedation-Radiation Interactive Effects*

Under sedation the *S*s in the previously irradiated group responded in some respects in a manner similar to that noted immediately following ionizing radiation. After irradiation *S*s showed a significant over-all increase in response time on the bent-wire detour problems with Life Saver rewards. A similar significant increase was noted for *S*s in this group on the most simple patterns, although there was a general trend of all *S*s, particularly the *S*s in the control group, to take less time to respond during the sedation period. Subjects in the experimental group failed more poker chip rewarded bent-wire detour problems during and after sedation than during the presedation period (see the third part of Figure 1). This initial rise is similar to that found during postradiation pathology, but the continued increase in failures is in contrast to the radiation study results.

4. *Radioantithetic Radiation-Sedation Interactive Effects*

The biggest difference between simple radiation effects and sedation induced radioantithetic effects on performance is the increased manipulation that occurs during the period of sedation contrasted with the lowered manipulation immediately following irradiation. The radioantithetic effects are exemplified by changes in the performance of previously irradiated *S*s on three different tasks requiring manipulation.

(*a*). During the period of sedation *S*s in the irradiated group manipulated the elements of the simple puzzles more than in the pre- or postsedation periods (refer to Figure 2). When the puzzles are presented to the *S*s in their living cages, all *S*s manipulate more elements during sedation than before or after the period of sedation. This effect was discussed along with other sedation effects but is exaggerated in the performance of the irradiated *Ss* (see Figure 2).

(*b*). The postsedation increase in failures by irradiation group *S*s on the bent-wire problem with poker chip reinforcement (this effect was noted as an exception in the discussion of the radiomemetic effects).

(*c*). Sedation appeared to inhibit the usual satiation of chewing on the wood block test and the inhibition is most manifest in the performance of the previously irradiated *S*s.

These changes indicate that irradiation effects are present over a long period of time. Subjects apparently can compensate for radiation effects but these effects are brought out under sedation debilitation.

E. Summary

Fifteen rhesus monkeys, nine previously irradiated with 400 *r* *X*-ray radiation and six nonirradiated control group *S*s, were trained on nine tasks before, during and after an 18-day period of chronic barbiturate sedation. Each *S* had received intensive prior training on the tasks in a radiation study ending four months before the present study began. Four classes of results were considered: (*a*) *S*s retention of tasks between prior training and the presedation period of the present experiment, (*b*) changes attributable to sedation *per se*, (*c*) interactive sedation-radiation changes that were radiomemetic, and (*d*) interactive sedation-radiation changes that were radioantithetic.

1. The *S*s' retention is very high for most of the tasks during the four months separating this and a prior experiment.
2. Changes in *S*s' performance resulting from sedation or sedation-radiation interaction were restricted to tasks requiring manipulation.
3. Changes in *S*s' performance that are attributable to sedation *per se* or sedation-radiation radiomemetic changes are less striking than the interactive changes that are radioantithetic.
4. Sedation and sedation-radiation interaction effects are radiomemetic in the respect of effecting *S*s' performance in the same tasks affected by irradiation.
5. The radioantithetic sedation-radiation interaction is exemplified by heightened manipulation during sedation, among previously irradiated *S*s, as contrasted with a decline in manipulation immediately following irradiation.
6. On the basis of the above findings *E*s do not find tenable the hypothesis of a singular behavior syndrome resulting from generalized sickness.

References

1. Cho, J. B., & Davis, R. T. Preferences of monkeys for non-food stimuli. *Amer. J. Psychol.*, 1957, **70**, 87-91.
2. Davis, R. T. Performance of rhesus monkeys on discrimination problems with reduced stimulus cues. *Proc. S. Dak. Acad. Sci.*, 1954, **33**, 116-122.
3. Davis, R. T., McDowell, A. A., Deter, C. W., & Steele, J. P. Performance of rhesus monkeys on selected laboratory tasks before and after a large single dose of whole-body *X*-radiation. *J. Comp. & Physiol. Psychol.*, 1956, **49**, 20-26.
4. Furchtgott, E. Behavioral effects of ionizing radiation. *Psychol. Bull.*, 1956, **53**, 321-334.

5. GRODSKY, M. A., CHO, J. B., & DAVIS, R. T. Paired comparison scaling of non-food stimuli. (Paper read at Amer. Psychol. Ass., New York, New York, 1954.)
6. HARLOW, H. F., DAVIS, R. T., SETTLAGE, P. H., & MEYER, D. R. Analysis of frontal and posterior association syndromes in brain damaged monkeys. *J. Comp. & Physiol. Psychol.*, 1952, **45**, 420-429.
7. McDOWELL, A. A., & DAVIS, R. T. A quantitative analysis of the chewing performance of rhesus monkeys. *Proc. S. Dak. Acad. Sci.*, 1954, **33**, 123-127.
8. McDOWELL, A. A., DAVIS, R. T., & STEELE, J. P. Application of direct observational methods to analysis of radiation syndrome in monkeys. *Percep. & Motor Skills*, 1956, **6**, 117-130.
9. SIEGEL, S. Nonparametric Statistics. New York: McGraw-Hill, 1956.
10. WEINER, R., & HARLOW, H. F. The effect of nembutal upon learned performances of the rhesus monkeys. *J. Gen. Psychol.*, 1952, **46**, 43-50.
11. WARREN, J. M., & HARLOW, H. F. Learned discrimination performance by monkeys after prolonged postoperative recovery from large cortical lesions. *J. Comp. & Physiol. Psychol.*, 1952, **45**, 120-126.

Department of Psychology
University of South Dakota
Vermillion, South Dakota

The Journal of Genetic Psychology, 1958, **93**, 53-57.

SEX, INTELLIGENCE, AND SORORITY OR FRATERNITY MEMBERSHIP AND CHANGES IN ETHNOCENTRISM OVER A TWO-YEAR PERIOD*

Department of Psychology, San Jose State College

WALTER T. PLANT

A. INTRODUCTION

In a recent study (2), it was found that there were significant changes in ethnocentrism associated with a two-year college experience. When subjects who had been enrolled in college for two years were compared with subjects like those enrolled, but who had not been enrolled in college for the two-year period, it was found that the college-enrolled subjects changed in the direction of decreased ethnocentrism. The non-enrolled subjects did not change in ethnocentrism.

It seemed desirable to determine what, if any, factors other than college enrollment were associated with a shift in ethnocentrism for the subjects who had completed the two-year college experience. An analysis of the original data permits a study of three such factors: sex and change in ethnocentrism, intelligence and change in ethnocentrism, and sorority or fraternity membership and change in ethnocentrism.

B. PROCEDURE

A 32-item, modified "Total Ethnocentrism Scale: Public Opinion Questionnaire E," developed by Levinson (1), was administered to each of 1,030 entering students at San Jose State College in the Spring of 1953. The same scale was administered to 755 of this original group in the Spring of 1955. Of this group of 755 who were retested, 505 had been enrolled at San Jose State College for the two academic years intervening between the two scale administrations, and this report is for the 505 subjects who had been college-enrolled.

For the study of sex and change in ethnocentrism, comparisons of mean 1953 *E*-scale scores and mean 1955 *E*-scale scores were made for 232 males and 273 females. To determine whether or not there were significant relationships between changes in ethnocentrism and intelligence, Pearson *r*'s were computed for ethnocentrism "shift scores" and raw scores on the 1949

---

*Received in the Editorial Office on May 6, 1957.

form of the *American Council on Education Psychological Examination* for the males and for the females. With regard to the study of sorority or fraternity membership and change in ethnocentrism, comparisons were made of mean 1953 *E*-scale scores and mean 1955 *E*-scale scores for sorority members and non-members and for fraternity members and non-members.

C. RESULTS

1. *Sex and Change in Ethnocentrism*

When the 1953 *E*-scale mean for the college-enrolled males who were retested two years later was compared with the 1953 *E*-scale mean for the college-enrolled females who were retested two years later, a significant difference between means was obtained. The mean score for females was significantly lower than that for males.

When mean 1955 *E*-scale score comparisons were made for these same groups, it was found that females had a significantly lower mean score than did males. Table 1 contains the summary data for these two comparisons.

TABLE 1

THE COMPARISON OF MALES AND FEMALES ON 1953 *E*-SCALE AND 1955 *E*-SCALE SCORES

	1953 *E*		1955 *E*	
	Males	Females	Males	Females
N	232	273	232	273
M	92.09	86.95	80.45	75.09
σ	24.39	25.88	22.38	24.90
t	2.29*		2.54*	

*Significant beyond the .05 level.

The obtained results seem quite clearly to indicate that females, who initiate and carry out a two-year college experience, were significantly less ethnocentric than males who initiate and carry out a two-year college experience, both at the time of entering college and after two years of time.

2. *Intelligence and Change in Ethnocentrism*

Shift scores were computed for each of the 505 subjects and these shift scores were correlated with raw scores on the 1949 form of the *ACE*. Shift scores were defined as the difference between 1953 *E* score and 1955 *E* score plus 100 (1955 E — 1953 E + 100). It was felt that such correlations would indicate the degree of relationship between change or shift in ethnocentrism over time and measured intelligence.

The obtained coefficient for females was —.12 and for males —.04. The coefficient for females is statistically significant while that for males is not.

The obtained coefficient for females does suggest a slight tendency for the brighter females to shift more in ethnocentrism than the less bright females in the sample, and the shift is in the direction of less ethnocentrism. The magnitude of the coefficient for females is sufficiently small though that when it is coupled with that for males, it can be interpreted to mean that change in ethnocentrism associated with a two-year college experience does not seem to be greater or lesser for individuals of varying intelligence.

3. *Sorority Membership and Change in Ethnocentrism*

It was found that 94 of the 273 females became members of a social sorority during their first two years in college. Statistical comparisons of the mean 1953 *E*-scale scores for sorority and non-sorority females indicated that sorority females were not significantly different from non-sorority females in ethnocentrism when they entered college. Comparisons of the mean 1955 *E*-scale scores for these same groups indicated that there was no significant difference in ethnocentrism between sorority and non-sorority females after a two-year college experience. Table 2 contains the summary data for these comparisons, and it will be seen that no significant differences prevailed between sorority and non-sorority females either when they entered college or after two years enrolled in college.

TABLE 2

THE COMPARISON OF SORORITY FEMALES AND NON-SORORITY FEMALES ON 1953 *E*-SCALE AND 1955 *E*-SCALE SCORES

	1953 *E*		1955 *E*	
	Sorority	Non-sorority	Sorority	Non-sorority
N	94	179	94	179
M	89.19	85.78	76.66	74.28
σ	28.35	24.44	23.79	25.40
t	.98*		.76*	

*Not significant.

4. *Fraternity Membership and Change in Ethnocentrism*

It was found that 62 of the 232 males became members of a social fraternity during their first two years in college. Statistical comparisons of the mean 1953 *E*-scale scores for fraternity and non-fraternity males indicated that fraternity males were not significantly different from non-fraternity males in ethnocentrism when they entered college. Comparisons of the mean 1955 *E*-scale scores for these same groups indicated that there was no significant difference in ethnocentrism between fraternity and non-fraternity males after a two-year college experience. Table 3 contains the summary data for these comparisons.

TABLE 3
THE COMPARISON OF FRATERNITY AND NON-FRATERNITY MALES ON 1953 *E*-SCALE AND 1955 *E*-SCALE SCORES

	1953 *E*		1955 *E*	
	Fraternity	Non-fraternity	Fraternity	Non-fraternity
N	65	167	65	167
M	88.04	93.66	83.42	79.29
σ	22.76	24.82	24.19	21.52
t	1.64*		1.20*	

*Not significant.

D. SUMMARY AND CONCLUSIONS

It has been demonstrated that a significant decrease in ethnocentrism was associated with a two-year college experience (2). The present study was based upon an analysis of data obtained in the cited study and was aimed at determining whether or not sex, intelligence, and sorority or fraternity membership were factors associated with changes in ethnocentrism among those enrolled in college during the two-year period.

It was found that:

1. Females enrolled in college for a two-year period were significantly less ethnocentric at the beginning of the two-year period and at the end of the two-year period than were males who were similarly enrolled.
2. An r of —.12 was obtained for a measure of shift in ethnocentrism over a two-year period and a measure of intelligence for the female sample of 273. An r of —.04 was obtained for the male sample of 232 on these same measures. The r for females was significant, but for males was not. In view of the small r for females and the non-significant r for males, it was concluded that changes in ethnocentrism associated with a two-year college experience are not greater or lesser for individuals of varying intelligence.
3. Sorority females were not significantly different in ethnocentrism when they entered college or at the end of two years of college from non-sorority females.
4. Fraternity males were not significantly different in ethnocentrism when they entered college or at the end of two years of college from non-fraternity males.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. The Authoritarian Personality. New York: Harper, 1950.
2. PLANT, W. T. Changes in ethnocentrism associated with a two-year college experience. *J. Genet. Psychol.*, (in press).

Department of Psychology
San Jose State College
San Jose, California

The Journal of Genetic Psychology, 1958, **93**, 59-62.

CHLORPROMAZINE AND HABIT REVERSAL*

Department of Psychology, University of Denver

ALFRED B. SHAKLEE[1]

A. INTRODUCTION

Evidence concerning possible effects of tranquilizing drugs on learning is as yet quite uncertain—partly because of difficulties in distinguishing between learning and other behaviors which may be drug sensitive. Several of the investigations to date have employed measures that might reflect motor, and other non-learning impairments, when the intention was to study such variables as learning, retention, or learned fear.

Chlorpromazine, in addition to its gross sedative effects, has been found to produce measurable changes in nervous system activity, and has in some cases been reported to influence learned behavior. Intravenous doses of 0.5-6.0 mg./kg. in rabbits have been shown to depress the *EEG* alerting reaction normally given after stimulation of the reticular formation (6). Study of the spontaneous electrical activity of the brain indicated (2) a greater proportion of slow frequencies from the motor cortex of the monkey following doses of 0.5-1.0 mg./kg., intramuscularly. This same research demonstrated that chlorpromazine in the monkey can reduce the threshold for after discharge following electrical stimulation of the amygdala and thalamus.

In a thirst-motivated, highly perfected maze habit in rats, no impairment in error score was noted following subcutaneous chlorpromazine injections of 0.5 and 10.0 mg./kg. (1). Running speed under large dose was depressed, but as the author emphasizes, this may have been due to lessened motivation. One study (4) has concluded that 10 mg./kg. subcutaneously in rats interfered slightly with the development of a conditioned emotional response, and clearly blocked experimental extinction. Other workers have indicated that an instrumental *CR* of climbing a pole to avoid shock was inhibited in rats by chlorpromazine, with minimal effects resulting from an oral dose of 5.0 mg./kg. (3).

For comparative purposes a recent report may be cited which suggested that human dosage for neurotic conditions has been fairly well standardized at approximately 0.7-7.0 mg./kg., daily (5).

---

*Received in the Editorial Office on May 20, 1957.

[1] This investigation was supported by a research grant (M1527) from the National Institute of Mental Health.

The present exploratory study has looked into possible effects of chlorpromazine on habit transfer and subsequent retention in the rat.

B. METHODS

Experimental animals were female hood rats between 31 and 38 days of age at the start of learning. The animals, with no previous history of experimentation, were assigned at random, five to a group, to receive respectively intraperitoneal injections of (*a*) 5.0 mg./kg. chlorpromazine, (*b*) 7.5 mg./kg. chlorpromazine, and (*c*) 0.2 ml. normal saline. The larger chlorpromazine dose had been suggested by pilot work to be about the greatest amount that would allow of fairly free running in the present maze. Both doses, however, produced noticeable sedation in the home cage situation.

A single unit *T*-maze, painted flat black, was used. The stem had inside dimensions of 6 x 10 in., and the arm, 6 x 32 in., with the two ends as goal boxes, each 6 x 12 in. Goal regions and the stem could be individually isolated by vertically sliding doors. The grid floor was of ⅛ in. diameter copper rods, placed ⅜ in. between centers. Hinged sections of plexiglass formed the top. Lighting was from 40-watt bulbs in reflectors centered 20 in. above each goal region, and a 60-watt bulb in a reflector 26 in. above the center of the stem. Motivation was produced by 0.35 ma current applied to the grid by a stimulator delivering rectangular, 60-cycle, direct current impulses, with magnitude quite independent of circuit resistance over a wide range.

Animals received taming experience during four days of the week prior to their respective first training trials. On both of the first two days, animals were each handled twice; on Day 3 they were allowed to explore the maze for 15 min. in groups of three to six; on Day 4 they were placed in the maze individually for 5 min. each.

Observations covered three training series for each animal—(*a*) learning a turning habit without drug, (*b*) learning the opposite habit under drug or control condition, and (*c*) relearning of the second habit 24 hrs. later. Errors were recorded for each of the three learning series, and latencies for the last two. Any movement of the fore part of the animal's body to the plane of the door of the wrong goal was counted as an error. On learning trials, the animal was placed in the stem, with all doors open and a charged circuit in the stem, choice area, and wrong goal. The stem was closed as the animal emerged into the choice region, but beyond this, retracing was allowed after an error. When the animal entered the correct goal region the door was closed. On the first two trials of initial learning, the animal was left in the correct goal for 1 min. before being removed. On all subsequent trials, the animal remained in the end box for approximately 10 sec. After a run, the animal was placed in an open-top box 24 in. in front of the stem and left there 45-60 sec. before the next trial.

Any time that an animal had remained motionless during a trial for 60 sec., it was given a light push with a stick in a direction perpendicular to the axis of the arm, or goal compartments.

On the first learning day, each animal was trained in a particular turning habit to a criterion of 10 out of 12 correct responses. For three animals in each group the left goal was unchanged; for two per group, the right goal. Immediately upon achieving the criterion an animal received his appropriate injection, and was replaced in an individual cage. After 2 hrs. to 2 hrs. 20 min., the second training series was begun, with the opposite turn now correct. The new learning was to a criterion of seven out of eight errorless trials. Approximately 24 hrs. later the animal received a "retention test" consisting of retraining on the last learned habit to the same criterion.

C. Results

As shown in Table 1, chlorpromazine was not demonstrated to affect acquisition of the new habit, nor its retention 24 hrs. later. While groups small as these could not lead to a statistical test with much power, the great similarity of the various group performances is of some special interest. Had a difference been demonstrated at the second or third learning series, it could have been ascribed to some combination of the two factors of drug, and shift in running conditions.

TABLE 1

Learning Measures in Chlorpromazine Comparison

Group	Trials to criterion: 1st Habit	Trials to criterion: 2nd Habit	Trials to criterion: 2nd Habit relearning	Shifts in latency: 2nd Habit	Shifts in latency: 2nd Habit relearning
Saline					
Mdn.	13	9	7	4	0.4
Range	12-20	8-12	7-11	1 to 11	—0.2 to 14.4
5.0 mg.					
Mdn.	14	12	8	5	1.0
Range	10-21	9-16	7-10	—4 to 9	—0.4 to 2.4
7.5. mg.					
Mdn.	11	11	7	—1	1.4
Range	10-14	8-18	8.14	—6 to 6	0.2 to 52.2

Pilot work had shown an apparently greater tendency for experimental animals to freeze on the early trials of drugged runs. This was the basis for the decision to record latencies in the second and third learning phases. A raw latency is an especially dangerous means of reflecting learning when there is possibility of differential motor impairment. For this reason a figure apparently less vulnerable to simple motor depression was employed. The median latency for the last three trials of a series was subtracted from the median of the first three to yield a measure of latency shift. These indices did not reflect differences in the present research, but seem worthy of study in other pharmaco-psychological investigation.

While one must emphasize the very great gap between certain human and lower mammalian behaviors, it is felt that the present type of design is capable of yielding some interesting starting points for inter-species generalization. As in some psychiatric tranquilizer applications, the present procedure is concerned with learning that took place in the absence of drugs, subsequent learning of a new pattern under drug influence, and later retention of the new habit when drug effect has been diminished.

The present finding that chlorpromazine prior to habit transfer influenced neither the new habit nor its retention may offer one useful landmark for more detailed future research.

References

1. Archer, J. D. Effects of chlorpromazine on maze performance of albino rats. *Fed. Proc.*, 1954, **13**, 332.
2. Delgado, J. M. R., & Mihailovic, L. Use of intracerebral electrodes to evaluate drugs that act on the central nervous system. *Ann. N. Y. Acad. Sci.*, 1956, **64**, 644-666.
3. Cook, L., & Weidley, E. Behavioral effects of some psychopharmacological agents. *Ann. N. Y. Acad. Sci.*, 1957, **66**, 740-752.
4. Hunt, H. F. Some effects of drugs on classical (Type *S*) conditioning. *Ann. N. Y. Acad. Sci.*, 1956, **65**, 258-267.
5. Margolis, L. H. Pharmacotherapy in psychiatry; a review. *Ann. N. Y. Acad. Sci.*, 1957, **66**, 698-717.
6. Rinaldi, F., & Himwich, H. E. Drugs affecting psychotic behavior and the function of the mesodiencephalic activating system. *Dis. Nerv. Syst.*, 1955, **16**, 133-141.

Department of Psychology
University of Denver
Denver 10, Colorado

The Journal of Genetic Psychology, 1958, **93**, 63-72.

COMPARISONS OF DISTRACTIBILITY IN IRRADIATED AND NONIRRADIATED MONKEYS*

School of Aviation Medicine, USAF

ARNOLD A. MCDOWELL[1]

A. PROBLEM

Two previous independent investigations have demonstrated significant changes in the free cage behavior of the rhesus monkey as a function of the radiation variable. Reid, *et al.* (7) observed changes during the irradiation period, while McDowell, *et al.* (6) showed similar changes during the period following irradiation. Hammack (3) also has demonstrated similar immediate posttreatment changes as a consequence of administering nitrogen mustard, a radiomimetic poison.

Scott (8), in his plea for greater emphasis on the technique of objective observation in the training of psychology students, points out that while experiments are conducted to test new ideas, they seldom result in new ideas. Discoveries come, he says, from the observation of a curious fact and the resultant inquiry into why this fact should happen.

Following Scott's dictum, the present study proposed to determine by systematic observation if selected categories of free cage behavior, which had shown change in the immediate posttreatment periods, were chronically changed. The study further proposed to test experimentally why such chronic changes should occur, if, in fact, they did occur.

B. METHODS

1. *Subjects*

Forty-three male rhesus (*Macaca mulatta*) monkeys were employed as subjects. These *S*s were part of another investigation and can best be grouped into three subgroups: a control group of 10 animals, a low dose group consisting of 18 animals, and a high dose group consisting of 15 animals. The dose range for the animals in the low dose group was between 60 and 160 rem, and the dose range received by the animals in the high

---

*Received in the Editorial Office on May 21, 1957.

[1] This work was conducted at the Radiobiological Laboratory of The University of Texas and the United States Air Force, Austin, Texas, and was supported (in part) by funds provided under Contract AF 18(600)-165 and the School of Aviation Medicine.

dose group ranged between 300 and 620 rem.[2] The source included both gamma and neutron radiation.

2. *General Methods*

A series of three experiments was conducted. The first experiment consisted of making frequency counts of selected items of behavior derived from the author's earlier studies of the cage behavior of monkeys (5). The second and third experiments were designed to describe the findings of the first experiment through more commonly employed laboratory instrumentation.

C. SPECIAL PROCEDURES AND RESULTS

1. *Systematic Observations of Free Cage Behavior*

a. Apparatus. The *S*s were observed in a special holding cage. This cage, supported at a height of 2 ft. from the floor, measured 2 by 2 by 2 ft. The back, two sides, and the floor of the cage were constructed of sheet aluminum. The ceiling was constructed of 1-in. heavy wire mesh. The front, which faced the observer, contained ¼-in. steel bars spaced 2 in. apart center-to-center. Approximately 1½ in. behind the back of the cage was a mirror which the animal could view through a 3-in. hole at a point 10 in. above the floor of the cage. A 6-in. piece of cotton rope clothesline and a 12- by 5-in. piece of cloth were in the cage during each observation.

b. Procedure. Selected categories of behavior were numbered on a multiple-selector switch which was wired into 23 counters. The observer watched the animal in the holding cage for 10 minutes and recorded (by depressing the switch over the appropriate numbered key) which of the 23 categories of behavior the animal was engaged in at the end of each 15-second period. Since the total observation period was of 10-minute duration, the total behavior category frequency per observation was 40. Each *S* was observed individually for four separate 10-minute periods during postirradiation months 11 and 12.

c. Results. Since the measurements were not normally distributed, White's rank test (1) was employed to determine the significance of difference between two groups.

In Figure 1 are shown the mean cage-directed category responses (visual, manual, or oral responses to cage parts) in each treatment group per successive observation. Statistical comparison of cage-directed category re-

---

[2] The dose ranges utilized represent a compromise of interests with departments in the Laboratory other than psychology. The subjects of this experiment, in consequence, were used because they were available and not because their radiation dosages were most appropriate to psychological research.

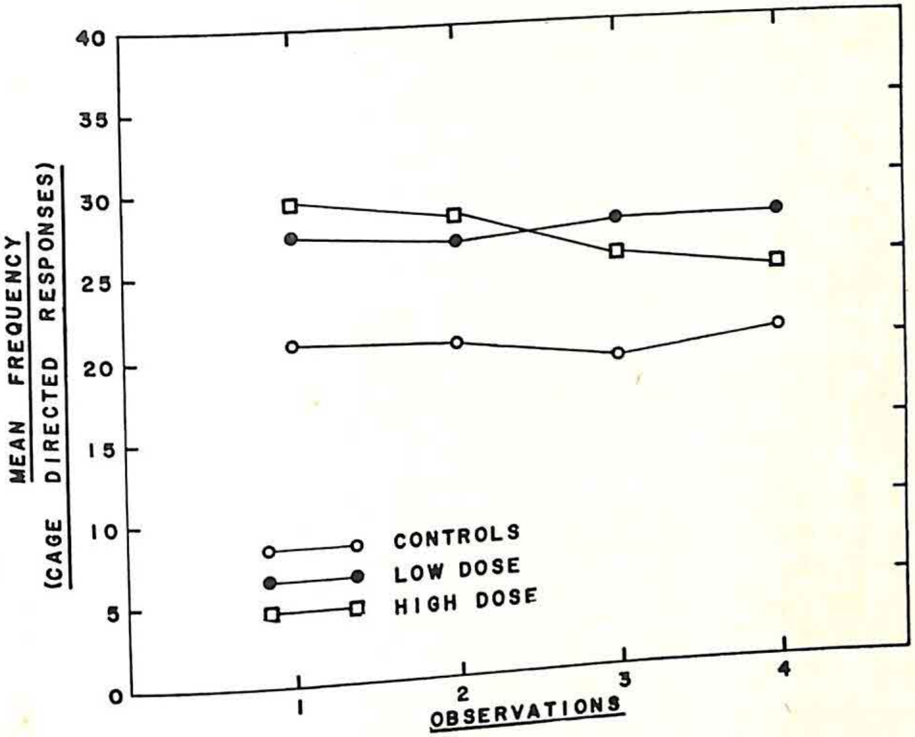

FIGURE 1
Mean Cage-Directed Responses for Control, Low-Dose Irradiated, and High-Dose Irradiated Rhesus Monkeys Per Successive Observation

sponses shows that (*a*) the *S*s in the control group had significantly fewer cage-directed responses than did *S*s in either the low or high dose experimental groups ($p = .05$ for both comparisons) and (*b*) the frequencies for the *S*s in the two experimental groups did not differ significantly.

Figure 2 compares the mean rapid energy expenditure responses (bouncing, pacing, and cage-shaking) for the *S*s in each group per successive observation. In contrast to the above finding, statistical comparisons with respect to rapid energy expenditure responses show that (*a*) the *S*s in the control group had significantly more rapid energy expenditure responses than did the *S*s in either the low or the high dose experimental groups ($p = .05$ for the comparison with the low dose group, and $p = .01$ for the comparison with the high dose group), and (*b*) the frequencies for the *S*s in the two experimental groups did not differ significantly.

The mean frequency of attention by the *S*s to uncontrolled auditory stimuli occurring outside the testroom (noises of other experimenters and other monkeys) is shown in Figure 3. Statistical treatment with respect to this

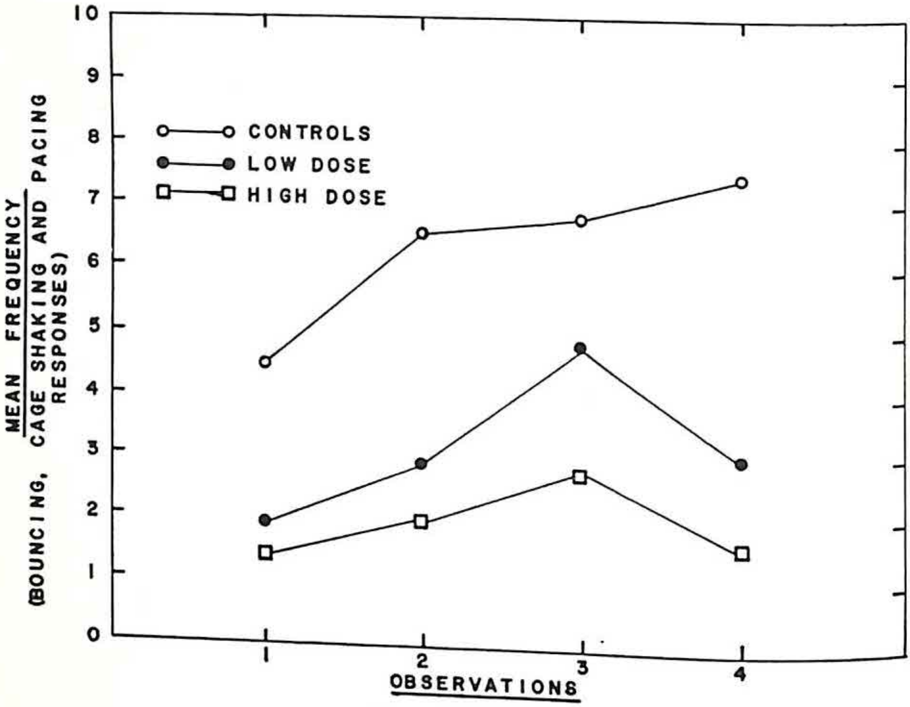

FIGURE 2

Mean Rapid Energy Expenditure Responses for Control, Low-Dose Irradiated, and High-Dose Irradiated Rhesus Monkeys Per Successive Observation

variable shows that (*a*) the *S*s in the control group did not differ significantly from the *S*s in the low dose experimental group, (*b*) the *S*s in the control group attended to extraneous auditory stimuli significantly more often than did *S*s in the high dose experimental group ($p = .05$), and (*c*) the *S*s in the two experimental groups did not differ significantly.

In consideration of the results on the category of attention to uncontrolled auditory stimuli occurring outside the testroom, it was hypothesized that with respect to extra-cage distractions (*a*) the control group would be most responsive, (*b*) the high dose experimental group would be least responsive, and (*c*) the low dose experimental group would define an intermediate position of response. It was further hypothesized, on the basis of the results concerned with the category of cage-directed responses, that all groups would be responsive to intra-cage distractions. To test these hypotheses, the following investigations were conducted.

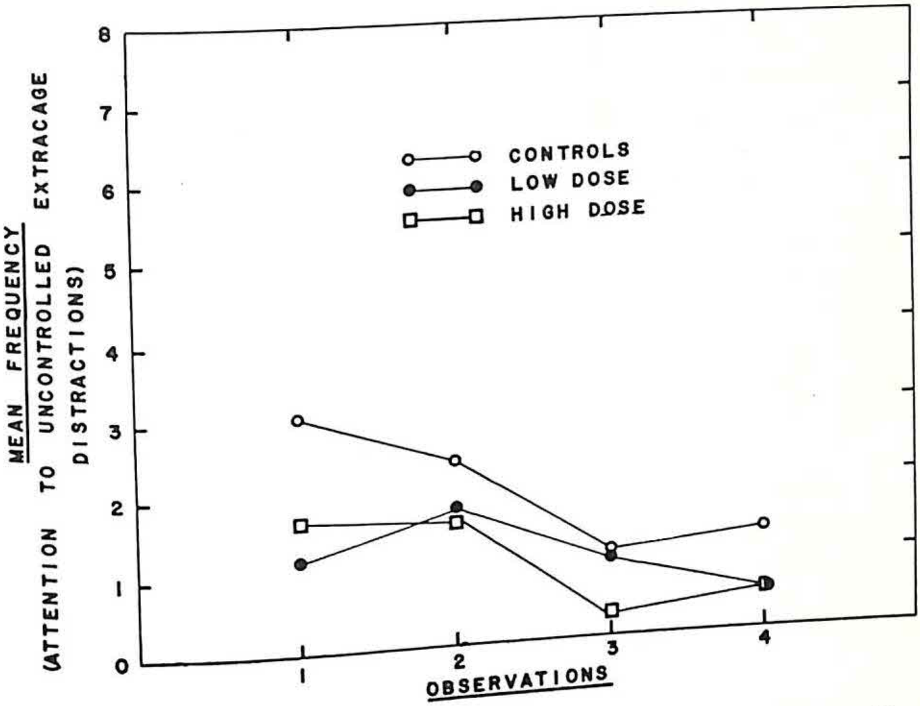

FIGURE 3

MEAN RESPONSES TO UNCONTROLLED EXTRA-CAGE DISTRACTIONS FOR CONTROL, LOW-DOSE IRRADIATED, AND HIGH-DOSE IRRADIATED MONKEYS PER SUCCESSIVE OBSERVATION

2. *Experimental Distraction*

a. General methods. The basic apparatus consisted of a *WGTA*, the test tray of which contained eight food wells, measuring 2 in. in diameter, and spaced 3 in. apart center-to-center. Each food well was covered by a 2 by 2 by 1/8 in. door, hinged in a fashion to open up and back. The entire test tray was painted gray.

A trial was run in the following manner: (*a*) each food well was baited with one-fourth of a grape and the food well doors were closed; (*b*) the tray was pushed forward to a full stop and the forward opaque screen was raised; (*c*) as soon as the opaque screen was raised a time clock was started; (*d*) when the *S* opened the door to the eighth food well the clock was stopped; (*e*) the tray was pulled back and the opaque screen lowered; (*f*) the experimenter recorded the performance latency and reset the clock; and (*g*) began the next trial.

In preliminary training, each *S* was tested seven trials on the first day, six trials on the second day, and two trials on each of the remaining six days. By the end of eight days of training, performance had become asymptotic.

b. Experiment I: Extra-cage stimulus conditions. Following the 8-day preliminary training period, a 9-day experiment concerned with extra-cage stimulus conditions was begun. Two distracting and one control condition were used in this experiment. An extra-cage *visual* distraction condition was arranged by affixing a 150-watt projector floodlight to the top of the test apparatus frame above the rear-view screen and by directing this flood at the location of the test tray in its forward position. The light flashed on and off intermittently, at the approximate rate of one second on and one second off, from the time the opaque screen was raised until the opaque screen was lowered. The extra-cage *auditory* distraction stimulus condition consisted of pounding with the handle end of a screwdriver on the bottom of the test table at the approximate rate of one beat every two seconds. Pounding commenced when the opaque screen was raised and ceased when the opaque screen was lowered. In this experiment each *S* was given two trials a day during each of the three 3-day periods. On one of the three days, a *S* worked under a control condition, on one day under an extra-cage visual distraction condition, and on one day under an extra-cage auditory distraction condition. The order in which *S*s were exposed to these conditions during each 3-day period was randomized.

c. Experiment II: Extra-cage plus intra-cage distraction conditions. After the 9-day training under extra-cage stimulus conditions, another 9-day experiment was conducted in a similar manner. This second experiment exposed each animal during each 3-day period to a normal condition, an extra-cage auditory distraction condition, and an intra-cage visual and auditory distraction condition.

The extra-cage auditory distraction condition was the playing of the tune "Good Night, Ladies" on a child's hand organ during the interval the opaque screen was raised. In the intra-cage distraction condition the rear panel door of the holding cage slid open immediately after the opaque screen was raised. The opening of this door produced a loud, scraping noise, and exposed a puzzle board containing 12 hooks and eyes, 12 loosely sunken eyebolts chained to the board, and 12 hasps.

d. Results. The coefficient of concordance (1) was used for the statistical treatment of the data. The statistical treatment was applied separately to the first trial and second trial data of each experiment. Ranks were assigned to the performances of each animal on the basis of longest performance latency, intermediate performance latency, and shortest performance latency.

Table 1 shows the coefficients of concordance for each of the three groups in Experiments I and II, computed for conditions within each 3-day session, and for each condition as a function of practice.

It can be seen from Table 1 that the animals in the control group on the first trials of Experiment I responded selectively to conditions within 3-day sessions throughout, and to both the light and sound stimulus conditions as a function of practice. The animals in the low dose experimental group on the first trials of Experiment I responded selectively to conditions within 3-day sessions for only the first 3-day session, and to the light stimulus condition as a function of practice. The animals in the high dose experimental group on the first trials of Experiment I responded in a chance manner throughout. Also, the performances of the animals in each group on the second trials of Experiment I are ordered in a chance manner throughout.

Table 1 also shows that the animals in the control group on the first trials of Experiment II responded selectively to conditions within 3-day sessions for the first and third 3-day sessions. The animals in the low dose experimental group on the first trials of Experiment II responded selectively to conditions within 3-day sessions throughout and to the combined intra-cage auditory and visual stimulus condition as a function of practice. The animals in the high dose experimental group on the first trials of Experiment II responded in a parallel manner to the animals in the low dose experimental group, with the single addition of selective response to the normal condition as a function of practice.

On the second trial of Experiment II, the animals in the control group responded selectively to conditions within 3-day sessions for the first 3-day session and to the combined auditory and visual stimulus condition as a function of practice. The animals in the low dose experimental group on the second trials of Experiment II responded selectively to conditions within 3-day session for the first and third 3-day sessions. The animals in the high dose experimental group in the second trials of Experiment II responded selectively to conditions within 3-day sessions for the first and second 3-day sessions.

D. Discussion

The results of this study are considered to support the hypothesis that irradiated *S*s show decreased distractibility, as Harlow and Moon (4) have recently suggested. The results further indicate that this decreased distractibility is limited to peripheral stimuli. Unpublished work by Grodsky (2), in which paired comparison preferences for different environmental

TABLE 1

COEFFICIENTS OF CONCORDANCE ON CONDITIONS WITHIN 3-DAY SESSIONS, SEQUENCE WITHIN 3-DAY SESSIONS, AND EACH CONDITION WITH PRACTICE FOR ALL GROUPS

Trial 1		Control group		Low-dose group		High-dose group	
		Exp. I	Exp. II	Exp. I	Exp. II	Exp. I	Exp. II
Conditions within 3-day sessions	Session 1	0.75**	0.43**	0.23*	0.38**	0.00	0.56**
	Session 2	0.47**	0.28	0.02	0.48**	0.05	0.59**
	Session 3	0.35*	0.37*	0.01	0.23*	0.05	0.23*
Sequence within 3-day sessions	Session 1	0.01	0.04	0.04	0.00	0.00	0.01
	Session 2	0.03	0.01	0.00	0.04	0.10	0.07
	Session 3	0.19	0.01	0.01	0.13	0.06	0.00
Conditions with practice	†Condition 1	0.00	0.27	0.07	0.04	0.08	0.21*
	†Condition 2	0.39*	0.28	0.13	0.00	0.02	0.12
	†Condition 3	0.49**	0.16	0.20*	0.50**	0.03	0.75**
Conditions within 3-day sessions	Session 1	0.19	0.79**	0.12	0.25*	0.08	0.50**
	Session 2	0.07	0.27	0.03	0.16	0.10	0.27**
	Session 3	0.16	0.13	0.02	0.26**	0.12	0.09
Sequence within 3-day sessions	Session 1	0.06	0.01	0.13	0.01	0.00	0.03
	Session 2	0.02	0.03	0.00	0.07	0.12	0.00
	Session 3	0.03	0.10	0.06	0.05	0.08	0.00
Conditions with practice	†Condition 1	0.13	0.01	0.18*	0.01	0.06	0.16
	†Condition 2	0.13	0.05	0.00	0.02	0.12	0.01
	†Condition 3	0.12	0.76**	0.06	0.09	0.09	0.14

*Significance at the .05 level.

**Significance at the .01 level.

†Condition 1 for both experiments is the normal condition. Condition 2 for both experiments is an extra-cage auditory stimulus condition. Condition 3 for Experiment I is an extra-cage visual stimulus condition; for Experiment II an intra-cage visual and auditory stimulus condition.

settings were scaled for control and irradiated monkeys, shows similar findings.

The differential responsiveness of the irradiated *S*s might best be described in Lewinian terms as a "reduction in life space." It is this change that probably produces the commonly reported superior performances of irradiated animals on laboratory tasks (4), since immediate stimuli are more readily dealt with by animals not distracted by peripheral stimuli.

E. Summary

The behavior of each of 10 rhesus monkeys serving as control *S*s and of each of 33 irradiated monkeys was systematically observed in a special holding cage over four 10-minute periods. On the basis of observed differences in behavior of the animals in the control and irradiated groups, it was hypothesized that the irradiated animals were less distractible than were the control animals. Two consecutive experiments were conducted to test this hypothesis. The results were as follows:

1. The control animals responded selectively to stimulus conditions occurring outside the confines of the cage, whereas the irradiated animals did so only transiently.
2. Both control and irradiated animals responded selectively to stimulus conditions occurring within the confines of the cage.
3. It was concluded that irradiated animals are less responsive to cage-peripheral stimuli than are control animals, but are just as responsive as control animals to cage-focal stimuli.

References

1. Edwards, A. L. Statistical Methods for the Behavioral Sciences. New York: Rinehart, 1954.
2. Grodsky, M. A. Preference scales of monkeys using visual and auditory stimuli. (Unpublished master's dissertation, Univ. S. Dakota, 1954.)
3. Hammack, B. A. Changes in the free-cage behavior of rhesus monkeys as a function of differential dosages of nitrogen mustard. (Paper read at Midwest Psychol. Assn., St. Louis, Mo., May, 1956.)
4. Harlow, H. F., & Moon, L. E. The effects of repeated doses of total-body *x*-radiation on motivation and learning in rhesus monkeys. *J. Comp. & Physiol. Psychol.*, 1956, **49**, 60-65.
5. McDowell, A. A. A normative study of the behavior of the caged macaque monkey. *Amer. Psychol.*, 1953, **8**, 395-396.
6. McDowell, A. A., Davis, R. T., & Steele, J. P. The application of systematic direct observational methods to the analysis of the radiation syndrome in monkeys. *Perceptual and Motor Skills*, 1956, **6**, 117-130.

7. Reid, J. B., Kaplan, S. J., & Melching, W. H. A method for the analysis of gross behavior. School of Aviation Medicine USAF Project Report No. 21-3501-0003, Report No. 10, June, 1955.

8. Scott, J. P. The place of observation in biological and psychological science. *Amer. Psychol.*, 1955, **10**, 61-64.

Radiobiological Laboratory
University of Texas
Austin, Texas

The Journal of Genetic Psychology, 1958, **93**, 73-78.

FACILITATIVE EFFECTS OF IRRADIATION ON PERFORMANCE OF MONKEYS ON DISCRIMINATION PROBLEMS WITH REDUCED STIMULUS CUES*

Department of Psychology, University of Texas

A. A. McDowell and W. Lynn Brown[1]

A. Problem

The hypothesis tested in the present study was derived from previous research by McDowell (7) concerned with comparisons of distractibility in chronically irradiated and nonirradiated rhesus monkeys. The results of this previous research demonstrated that irradiated animals are less responsive to cage-peripheral stimuli than are control animals, but are just as responsive as control animals to cage-focal stimuli.

On the basis of this demonstrated difference in distractibility between irradiated and control animals, it was hypothesized that irradiated animals would eliminate errors faster than control animals on discrimination problems with reduced stimulus cues (3, 4).

B. Methods

1. *Subjects*

Twenty-four male rhesus (*Macaca mulatta*) monkeys were employed as subjects. These *S*s were part of another investigation and can best be grouped into three subgroups: a control group of nine animals, a low-dose group consisting of eight animals, and a high-dose group consisting of seven animals. The dose range for the animals in the low-dose group was between 10 to 16 neutron rep plus 70 to 140 gamma roentgens, and the dose range received by the animals in the high-dose group was between 27 to 54 neutron rep plus 284 to 557 gamma roentgens, the source including both gamma and neutron radiation.

2. *Apparatus*

All training of animals in this experiment was conducted in a modified version of the Wisconsin General Test Apparatus (*WGTA*), the holding

---

*Received in the Editorial Office on May 21, 1957.

[1] This work was conducted at the Radiobiological Laboratory of The University of Texas and the United States Air Force, Austin, Texas, and was supported (in part) with funds provided under Contract AF 41(657)-149 and the School of Aviation Medicine.

cage of which measured 36 by 30 by 24 inches. A stimulus tray holding two foodwells, spaced 12 inches apart center-to-center, was used. The foodwells were each covered with identical red painted square wooden blocks to prevent the animal's seeing the food reward. Each of the square wooden blocks measured 4 by 4 by ¾ inches. An added stimulus cue object consisted of a white painted round wooden block measuring 3⅜ by ¾ inches.

3. *Procedure*

A reduced cue discrimination trial was presented in the following manner. The opaque screen of the *WGTA* was lowered ending the previous trial and the stimulus tray was cleared. One raisin was placed in either the right or left side foodwell according to a predetermined random order. The identical square wooden blocks were placed simultaneously over both foodwells. The round wooden block was then placed on the square wooden block covering the food reward. The one-way vision screen was lowered, the stimulus tray carrier advanced, and the opaque screen raised allowing the animal to make a choice.

After the animal completed the first trial, the stimulus tray carrier was withdrawn from the reach of the monkey, the opaque screen lowered, and the second trial commenced. The monkey was allowed only one choice on each trial. The second trial was similar to the first with the exception that the round wooden block was not used as a cue and the correct (rewarded) response was to the square wooden block which was rewarded on the first trial of the problem.

Each animal was tested on 24 problems per day for 24 days, a total of 576 reduced cue discrimination problems. The order of occurrence of randomly and constantly arranged blocks was randomized in the 24 problems presented to each monkey each day.

The number of errors made by each monkey on the second trial of each problem was recorded and the results were arranged for analysis. Because the assumption of homogeneity of within-group variances could not be met, the Mann-Whitney *U*-test (9) was used for the statistical analysis.

C. RESULTS

Figure 1 shows the per cent errors for each group on successive 6-day testing periods. The frequency of total errors for the animals of the high-dose group is significantly less than the frequency of total errors for the animals of the control group at the 5 per cent level of confidence, and is less than the frequency of total errors for the animals in the low-dose group at the

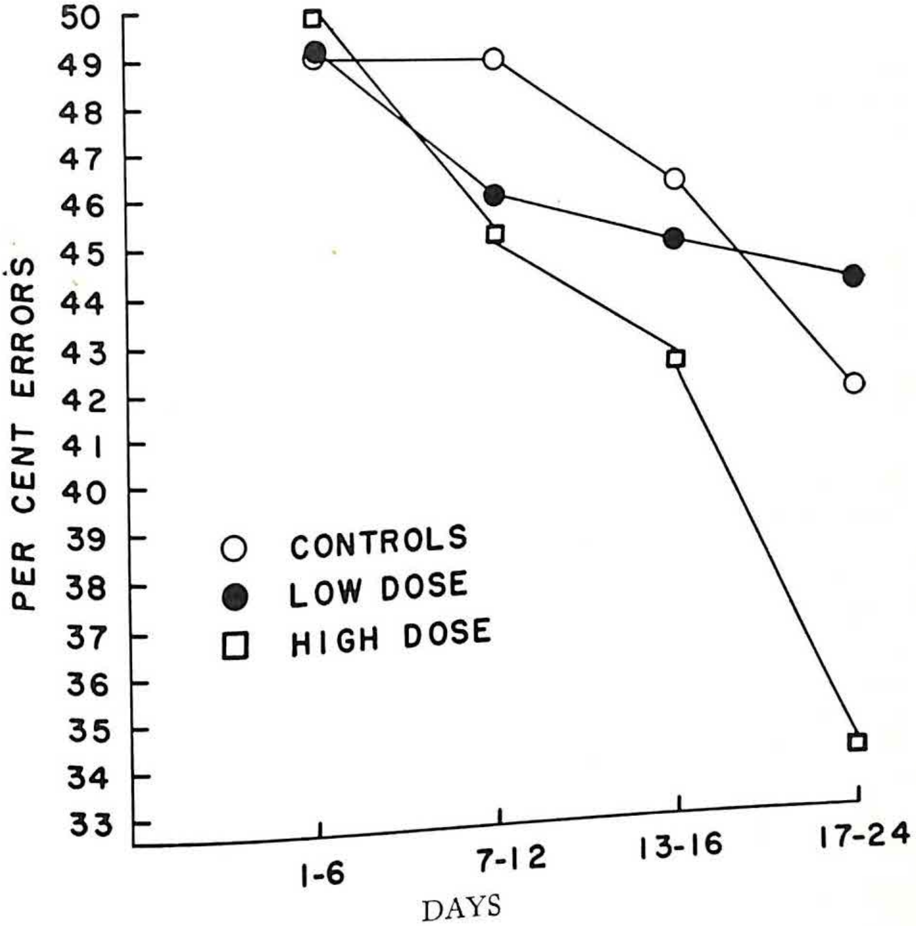

FIGURE 1

Facilitated Reduced Cue Discrimination Performance of High-Dose Irradiated Monkeys

6 per cent level of confidence. The frequency of errors for the animals of the high-dose group during Days 13 through 18 and during Days 19 through 24 is significantly less at the 5 per cent confidence level than the frequency of errors for the animals in the control group during the same two periods. The frequency of errors for the animals of the high-dose group during Days 19 through 24 is significantly less at the 2 per cent confidence level than the frequency of errors for the animals of the low-dose group during the same period. No other statistical comparisons yielded significant results.

Although the groups were formed on the basis of random selection prior to the introduction of the radiation variable, the possibility exists that some of the group differences may be due to prior differences in the animals' abilities to learn the task presented to them. The animals were given preirradiation training on delayed response problems. Table 1 presents the mean errors per 25 trials per group per day over 50 days of training on this preirradiation testing, and the correlations between errors on the preirradiation delayed response problem and the postirradiation errors on the discrimination problems with reduced stimulus cues. There were no significant group differences on the preirradiation delayed response problem. Errors on this problem correlated significantly with errors on the problem of the present study for the control group. Similar correlations for the low- and high-dose goups were not significant.

TABLE 1
MEAN ERRORS PER DAY ON PREIRRADIATION DELAYED RESPONSE TRAINING FOR EACH GROUP AND RANK ORDER CORRELATION COEFFICIENTS FOR EACH GROUP CORRELATING PREIRRADIATION DELAYED RESPONSE ERRORS AND POSTIRRADIATION REDUCED CUE ERRORS

	Mean errors	Rho
Controls	8.25	.63*
Low-dose	8.01	.25
High-dose	8.58	—.29

*Denotes significance beyond the 5 per cent confidence level.

D. DISCUSSION

The results of the present study support the hypothesis that irradiation will facilitate the performance of monkeys on discrimination problems with reduced stimulus cues. These findings are in accord with the suggestions of Harlow and Moon (5), and of Riopelle, Grodsky, and Ades (8), working with monkeys, and of Blair and Arnold (1) working with rats, that irradiation may facilitate learning task performance. It is probable that performance facilitation in the irradiated monkey is a direct resultant of chronically decreased responsiveness to cage-peripheral stimuli.

The concept of facilitation of learning performance by debilitation is not new. Lashley (6) in 1920 reported that initial learning of a light-dark discrimination by partially decorticate rats was superior to that of normal animals, both in terms of trials and errors. In 1937, Brogden and Culler (2), using conditioned response technics to measure absolute loudness thresholds, reported a transitory gain in acoustic sensitivity of dogs following *X*-irradiation. More recently Wade (11) has demonstrated that prefrontal monkeys under Dial or Nembutal administration manifest the ability to perform above chance level on delayed response problems.

Stanley and Jaynes (10), in a consideration of Wade's results, conclude that the physiological effect of sedation is to raise the thresholds of all the responses to the many stimuli constantly surrounding the animal to the same degree. In consequence, the response provoking potentialities of the weaker stimuli are lost and only those responses with the greatest habit strengths are provoked; i.e., the well-learned responses to the baited foodwells.

McDowell (7) has previously demonstrated the same course of events in the irradiated animal. This decreased responsiveness to the weaker stimuli in the irradiated monkey is reflected, as the present study shows, in facilitation of learning performance on the delayed response type problem just as is the sedation effect in the prefrontal monkey. It appears, however, that the point of departure for facilitation in the irradiated animal is from the response threshold values of the normal animal rather than from the lowered threshold values of the prefrontal animal. It is concluded that irradiation, with respect to alteration of response thresholds, is sedationmimetic, and that irradiation, in a manner as yet unspecified, acts as a chronic, mild sedative.

E. SUMMARY

Nine control, eight low-dose irradiated, and seven high-dose irradiated monkeys were used to test the hypothesis that chronic irradiation will facilitate the learning performance of monkeys on discrimination problems with reduced stimulus cues. The results of the experiment supported the hypothesis. These results are discussed as suggesting that chronic irradiation, with respect to effect upon response threshold values, acts as a chronic, mild sedative.

REFERENCES

1. BLAIR, W. C., & ARNOLD, W. J. The effects of cranial X-radiation on retention of maze learning in rats. *J. Comp. & Physiol. Psychol.*, 1956, **49**, 525-528.
2. BROGDEN, W. J., & CULLER, E. Increased acoustic sensitivity in dogs following roentgen-radiation of the hypophysis. *Amer. J. Physiol.*, 1937, **119**, 13-23.
3. COWLES, A. T., & NISSEN, H. W. Reward-expectancy in delayed response of chimpanzees. *J. Comp. Psychol.*, 1937, **24**, 345-358.
4. DAVIS, R. T. The performance of rhesus monkeys on discrimination problems with reduced stimulus cues. *Proc. S. Dak. Acad. Sci.*, 1954, **33**, 1-7.
5. HARLOW, H. F., & MOON, T. E. The effects of repeated doses of total-body X-radiation on motivation and learning in rhesus monkeys. *J. Comp. & Physiol. Psychol.*, 1956, **49**, 65.
6. LASHLEY, K. S. Cerebral function in learning. *Psychobiology*, 1920, **2**, 55-136.
7. MCDOWELL, A. A. Comparisons of distractibility in irradiated and nonirradiated monkeys. *J. Genet. Psychol.*, 1958, **93**, 63-72.

8. RIOPELLE, A. J., GRODSKY, M. A., & ADES, H. W. Learned performance of monkeys after single and repeated *X*-irradiation. *J. Comp. & Physiol. Psychol.*, 1956, **49**, 521-524.
9. SIEGEL, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.
10. STANLEY, W. C., & JAYNES, J. The function of the frontal cortex. *Psychol. Rev.*, 1949, **56**, 18-31.
11. WADE, M. The effect of sedatives upon delayed response in monkeys following removal of the prefrontal lobes. *J. Neurophysiol.*, 1947, **10**, 57-61.

Department of Psychology
University of Texas
Austin, Texas

The Journal of Genetic Psychology, 1958, **93**, 79-85.

REACTIONS OF SUPERIOR AND FEEBLEMINDED CHILDREN TO AN ILLUSION*

Department of Psychology, University of Florida

J. C. DIXON[1]

A. PROBLEM

This study compares how superior and feebleminded children react to an illusion which poses a conflict between perception and expectation. The illusion presented to the child is the ring-segments illusion. The two segments are equal in size, but when placed one above the other in the same plane, the lower segment appears distinctly larger. Reversal of the segments points up the illusion and provides an incongruous situation in which the previously larger-appearing segment suddenly appears smaller and the smaller larger.

In an earlier study with nursery and kindergarten children (1) it was observed that some resolved their conflict by continuing to choose as larger the segment which was consistent with their last choice but apparently contrary to their present perception, while others seemed to choose in accordance with their present perception. (For convenience, the former kind of choice will be called *E*, for expectancy, and choice in accord with present perception will be called *P*.) There was a tendency for the older children to make more *P* choices than the younger. Later informal testing with older grade school children also suggested a marked *P* preference, but *MA* was not controlled in either instance so it is not known whether this preference is related more to *CA* or to *MA*.

It goes without saying that a certain amount and kind of experience are necessary before one can appreciate a discrepancy. The naïve child has little experience to contradict his eyes, while the dull-witted has little ability to make effective use of his experience. How the clash between expectation and contradictory experience is resolved no doubt depends on a number of other factors too, but a beginning may be made by relating this choice behavior to *MA* as an index of mental maturity and to *CA* as an index of amount of experience.

---

*Received in the Editorial Office on June 3, 1957.

[1] This investigation was supported by a research grant M-1326 from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

B. METHOD

1. *Subjects*

A group of 30 superior nursery school children and 60 feebleminded children, equated for *MA*, were the subjects. Each of these two groups was divided into three equal groups on the basis of Stanford-Binet *MA*, about 4, 5, and 6 respectively, to find out if there was a shift from *E* to *P* in this *MA* range. The feebleminded group was further divided into a *young* and *old* group of the same *MA* to study the relation between age and choice behavior within that group.

Table 1 shows the mean and standard deviation of *CA* and *MA* in months for the three *MA* groups. It will be noted that each *MA* grouping of 30 contains three subgroups of 10 each: superior, young, and old feebleminded.

TABLE 1
MA AND *CA*, IN MONTHS, OF THREE *MA* GROUPS

MA Group	*N*	*MA* Mean	*MA* *SD*	*CA* Mean	*CA* *SD*
Four					
Superior	10	47	6	38	2
Feebleminded					
Young	10	48	4	125	11
Old	10	48	3	192	24
Five					
Superior	10	60	7	43	3
Feebleminded					
Young	10	60	2	134	13
Old	10	59	2	204	25
Six					
Superior	10	70	6	60	4
Feebleminded					
Young	10	73	5	142	9
Old	10	73	4	175	12

2. *Procedure*

Before being presented with the ring-segments, the child was tested with four pairs of non-illusory figures, large and small stars, circles, squares, and rectangles to be sure he could tell big from little, and if he made a mistake, he was excused from completing the experiment. Each member of a pair of figures was differently colored to aid in identification. The ring-segments, one red and one blue, were then presented as part of the series and, as with the other figures, he was first asked to point out the big one. Then the segments were reversed, care being taken that the child was observing the process, and he was asked again to point out the big one. This is the crucial response. When expectation and perception are in conflict, which segment does he choose? Does he make an *E* choice, i.e., select the segment which is consistent with his remembered last choice (but which contradicts his present perception), or does he make a *P* choice, i.e., choose in accordance with his present perception (but inconsistent with his last choice)?

As a measure of consistency of choice, the illusion was again presented and his response noted. It is therefore possible for the child to choose on a *P* basis one time and *E* the other. These cases will be called *M*, for *mixed*. To avoid color memory effects, yellow and green segments were substituted for the red and blue. Further inquiry was then made to determine how he resolved the incongruous situation.

C. RESULTS

When the ring-segments were reversed most of the four-year *MA* group, both superior and feebleminded, made consistent *E* choices. At the five-year level most of the feebleminded children continued to make consistent *E* choices but the superior children began to shift toward *P*, showing a preponderance of *M* (mixed) choices. At six the superior children have completed the shift to *P* while the feebleminded seem to be in transition between *E* and *P*. The choices of these two groups are compared by *MA* levels in Table 2.

The superior and feebleminded groups were combined and a chi-square

TABLE 2
CHOICES OF THREE *MA* GROUPS OF SUPERIOR AND FEEBLEMINDED CHILDREN

MA Group	*N*	Choice*		
		E	*M*	*P*
Four				
Superior	10	5	3	2
FM Young	10	7	2	1
Old	10	6	1	3
Five				
Superior	10	2	5	3
FM Young	10	6	3	1
Old	10	8	1	1
Six				
Superior	10	1	2	7
FM Young	10	4	1	5
Old	10	4	2	4

**E*: choice in terms of expectancy; *P*: choice in terms of present perception; *M*: mixed.

test of the differences among the three *MA* groups in distribution of *E*, *M*, and *P* choices was made. A chi-square value of 12.65, with four degrees of freedom, yields a *p* between .02 and .01 thus suggesting that *P* choices vary with *MA*. As noted above, the shift from *E* to *P* is greater in the superior group. This difference between the two groups on choice distribution is significant (chi-square of 8.20, with two degrees of freedom, yields a *p* between .02 and .01).

To test further the relation between choice and *MA*, an analysis of variance was run separately for the two groups. For the superior group, *F* is 3.56 which, with two and 27 degrees of freedom, is significant at better than the .05 level. For the feebleminded group, *F* is 2.60 which, with two and 57 degrees of freedom, is not significant at the .05 level. Thus, the differences are likely due to the greater shift in choice in the superior group.

A comparison of the *young* and *old* feebleminded, equated for *MA*, was made to find out if *CA* was associated with change in choice. The resulting chi-square of .87, with two degrees of freedom, yields a *p* between .70 and .50. Thus, choice does not vary as a function of *CA* in this group.

It is fairly clear that the differences and trends both within the feebleminded groups and between the superior and feebleminded groups vary with *MA* more than *CA*. What the shift from *E* to *P* means is another question.

One possibility is that, by responding first to the non-illusory figures, the child is being conditioned to give a place response. It will be recalled that the non-illusory figures, the stars, circles, squares, and rectangles, were reversed exactly like the ring-segments and that the ring-segments were last in the series. The child might then develop a set to pick up the *reversed* figure as big after reversal since it is the larger figure which is presented initially as the lower figure and then reversed.

To find out actually what effect the non-illusory figures might have had on choice, the ring-segments were presented *before*, rather than after the non-illusory figures, to 20 additional feebleminded and 20 superior children whose mean mental ages were about 5½ years (64 and 66 months respectively). The standard deviations of these groups are 7 and 9 months respectively. The feebleminded were between the ages of 16 and 20 years, mean 18.2, standard deviation four years. The superior children were between 36 and 66 months, mean 54, standard deviation, seven months.

The changed procedure was highly effective in changing segment choice. Table 3 compares the superior and feebleminded children on the choice categories under the new procedure (*B*), and under the former procedure (*A*).

The differences between procedures, for both groups, yield chi-square values significant beyond the .01 level (comparison of the superior groups yielded a chi-square value of 34; comparison of the feebleminded groups, 20.57). The differences between the superior and feebleminded groups under the changed procedure are not significant, however. A chi-square value of 3.39, with two degrees of freedom, is not significant at the .10 level.

TABLE 3
COMPARISON OF TWO GROUPS UNDER PROCEDURES *A* AND *B*

Group	Procedure	Choice *E*	 *M*	 *P*
Superior	A	8	10	12
Feebleminded	A	35	10	15
Superior	B	0	2	18
Feebleminded	B	3	4	13

Since the choice differences between the two groups are significant under procedure *A*, and not under procedure *B*; and since the chief difference between the two procedures appears to be that procedure *A* fosters a set or expectation to make *E* choices, it seems reasonable to conclude that the feebleminded children are more easily influenced by the set-inducing procedure *A*. It would be less accurate, however, to say that they more readily developed a response set than to say that they less readily responded adaptively. That is, their response set was not necessarily more easily developed but more resistant to change. The concept of cognitive rigidity (3) may also be useful in conceptualizing this behavior in which communication between memory or expectation and perception is not effective.

D. ADDENDUM

It may be of incidental interest to report something of the children's comments when asked to explain the unusual phenomenon they are witnessing. Each child, after completing his choices, was invited to manipulate the segments himself and, as he examined and reversed the segments, an inquiry was made, e.g., "It looks big and then it looks little. I wonder how come? Tell me, what makes it look big and then little?"

Comments ranged from *none* or "I don't know," to *descriptive* comments such as, "It looked like it was little, then it got big," to *attempted explanations* such as, "You swap them around" or "I bet it's the way they're cut." Table 4 shows the number of children (procedures *A* and *B* combined) responding in these ways to the inquiry.

About a third of the superior and a quarter of the feebleminded attempted

TABLE 4
RESPONSES TO INQUIRY MADE BY THREE *MA* GROUPS OF SUPERIOR AND FEEBLEMINDED CHILDREN

MA Group	None	Response Description	Explanation
Four			
Superior	10	3	0
FM	21	0	2
Five			
Superior	5	5	5
FM	14	6	9
Six			
Superior	5	4	13
FM	12	4	12

explanations. At *MA* 4 the children of both groups typically respond with "I don't know" or unverbalized amusement and wonder, but the 5-year-olds venture both descriptive comments and explanations, while the 6-year-olds conform best to the request to attribute the change to some cause. The differences between the superior and feebleminded are not significant (a chi-square of 4.89, with two degrees of freedom, gives a P between .10 and .05).

The number of explanations is too few to attempt a classification of kinds of explanation by *MA* groups. A majority of both the superior and feebleminded, 12 in each group, attempted to explain the illusion in terms of the conditions under which it was produced, e.g., "Every time you put it on bottom it gets big, on top it gets little." The other comments were distributed as follows: five superior and one feebleminded called it a magical trick or puzzle, one superior and four feebleminded attributed the change to the shape of the segments, two feebleminded said simply, "You did it," and one each replied, "On account of it growed," "The colors," "It wants to."

Huang (2), whose study of children's explanations of the Müller-Lyer illusion most closely approximates this aspect of the present study, also reported that his children most frequently attempted to explain the illusion in terms of the circumstances observed to accompany it rather than by animistic or magical kinds of explanations.

F. SUMMARY

Feebleminded and superior children, equated for *MA*, responded to the ring-segments illusion which poses a conflict between perception and memory or expectation. Resolution of the conflict by choosing the segment as larger which presently appeared larger, rather than the one which had appeared larger before, was more characteristic of increasing mental maturity than chronological age. Choice was also found to be highly dependent on the set-inducing nature of the procedure regardless of MA or CA, and more so for the feebleminded than for the superior children.

REFERENCES

1. DIXON, J. C. Concept formation and emergence of contradictory relations. *J. Exp. Psychol.*, 1949, **39**, 144-149.
2. HUANG, I. Children's explanations of strange phenomena. *Psychol. Forsch.*, 1930, **14**, 63-182.
3. WERNER, H. The concept of rigidity; a critical evaluation. *Psychol. Rev.*, 1946, **53**, 43-52.

Department of Psychology
University of Florida
Gainesville, Florida

The Journal of Genetic Psychology, 1958, **93**, 87-111.

A STUDY OF SOCIALIZATION PATTERNS AT THE HIGH SCHOOL LEVEL*

Department of Education and Institute of Child Welfare, University of California

Mary Cover Jones

A. Problem

The school and the peer group as socializing agents are interdependent. With the growth of mass education in our culture and the use of education as a social ladder, schools have become more and more aware of the importance of this interaction as a motivating force in learning. One evidence of this is the increasing value placed upon extracurricular functions as part of the school program. The collaboration of the school and the peer group in promoting activities common to their social goals is a powerful and pervasive force in the high school student's daily life. Some students respond to this socializing pressure and identify with its purposes. Others resist or escape its influence in varying degrees.

Can we specify some of the factors—pertaining perhaps to intelligence, physical maturity, social class status, or social attitudes—which may make individuals psychologically more or less accessible to the molding influences of the school through its program of social participation? What criteria can we use to determine which students are responding to, which evading, the efforts at socialization?

B. Procedure

In an attempt to answer some of these questions, the present study contrasts two groups of high school students (members of the Adolescent Growth Study of the Institute of Child Welfare) who fall at opposed extremes of a sample distributed on the basis of frequency of mention in the daily newspaper of the high school which they attended for three years. The analysis may be looked upon as selecting individuals according to the rôles which they play in one area of their social living, that of contributing to group life in high school.[1] On the one hand we have the students who identified themselves with the purposes and plans of this large amorphous

---

*Received in the Editorial Office on June 5, 1957.

[1] "It's easy to classify people by skin color, sex, or age; it's relatively easy to classify them by the number of dollars they earn in a year. Very little grouping of human beings is done in terms of the rôles they reveal in their own immediate social world." Murphy, G. (12, p. 23).

population (the student body), and who were successful in demonstrating this identification to the faculty and to their peers. On the other hand, there are those students who received no public recognition (in the school news organ) because they lacked either the ability or the inclination to be drawn into activities which were a matter of common interest. They were so little identified with the school subculture as to ignore or deny the pressures inherent in its value systems.

The newspaper which served as the basis of selection for these two opposed groups was the daily publication of a co-educational, public senior high school in a large western city. The school also served as a teacher training center for student teachers from the education department of a state university and was one of the coöperating schools in a nation wide educational research program (4). The administrative and teaching staff enunciated and fostered the concept of the school as a broadly functioning institution: "In contrast to the traditional emphasis upon preparation for living, University High School holds that adolescents are living here and now and that it is the school's business to help them to live successfully" (4, pp. 68-69).

The school had prestige as an educational and social institution; it tended to attract able students of superior home backgrounds (largely upper middle class) from all areas of the city and from neighboring communities.[2] It was also the neighborhood school for a large proportion of those who attended. For many of the latter, with an admixture of lower middle class families, it was simply the school most conveniently located for them.

Whether or not the students had selected the school because of its status as an educational institution they were early and often reminded of its unique atmosphere. Because this school served as a pilot for educational policy and research, it was articulate in defining its tenets and active in motivating staff and students toward a recognized goal.[3] One approach to its educational aims was through enlisting the peer group in extensive activities of the sort usually classed as extracurricular. That this aspect of school life was lively and successful is attested by the fact that at the time of one report, 86 per cent of the student body engaged in some form of extra-class school activity (3).

---

[2] About 20 per cent of the student body were from outside of the school district and represented those families which sought admittance to the school because of its educational and social advantages. None of this 20 per cent (a relatively high status group from the neighboring university city) were included in the study.

[3] In a published discussion of the school's basic philosophy this statement was made: "Fortunately because this philosophy was clearly formulated . . . the school was established upon a system of enduring values and was developing according to a well-defined but continuously evolving plan" (4, p. 67).

The daily newspaper (one of the extracurricular activities) was devoted to the reporting of school affairs exclusively. It therefore served much as an official "house organ," reflecting the interests, activities, and values of the student group. It instructed its readers as to where and when they could buy tickets for events, what to see in exhibit cases, how to sign up for tryouts, what elective and appointive positions were available, and, in general, what the extracurricular program offered and who was prominent in school affairs. Editorials, signed articles which reviewed events or evaluated activities, and letters to the editor were also included.

Occasionally a student received recognition for an achievement or honor attained under other than school auspices such as appearance on a radio program or selection as a delegate to a Scout conference. Those who were never mentioned in even such a passing manner as "credit goes to the following students who have been associated with office mimeographing . . ." or "among those who also tried out for parts were . . ." must, in general, have found school a narrower and less socially molding experience than those who participated in school life to an extent which brought frequent recognition. Participation in extracurricular activities was the accepted and approved pattern and a certain amount of pressure was exerted through the daily paper to foster this kind of behavior. "Girls' League President Asks for Girls' Support of Afternoon Teas" . . . "Dean of Girls Asks Members of League to Support Vocational Talks" . . . Such headlines as these appeared frequently in the daily newssheets.

Five hundred and forty issues (covering a three-year period) were read and all mentions tabulated for the 122 study members who attended this high school throughout the 10th, 11th, and 12th grades. The total number of mentions for different individuals ranged from zero to 112. The zero category included 9 boys and 9 girls, and these were selected as representing the "low prominence" or "no-mention" group. In selecting a group at the high-mention extreme, consideration was given to the differing significance of items; it was decided that a weighted score (based on prestige) would provide a more defensible basis for determining prominence than would a raw frequency score. Table 1 illustrates the categories into which the items were classified, and the assigned weights, on an estimated prestige scale.

The weighting of news items was derived from the composite judgment of staff members who were intimately acquainted with the school. This introduces a subjective factor which was uncontrolled by the use of formal scaling methods. However, in selecting for further study the 12 cases of

TABLE 1

DESCRIPTIVE CATEGORIES FOR CLASSIFYING NAMES MENTIONED

Category	Example	Assigned weight
I. *School leadership*		
A. Student body president.	I. "*Karl Envick*, new student-body president, gives welcome to new pupils."	15
B. Chairman of major committees or school groups.	"So congratulations to *Sheila Harmon* for putting on a swell Follies. . . ."	12
C. Member of major committee or council, student body officer, president or chairman of minor committee.	"'I want to welcome all L-10 boys to this school,' stated *Bob Rank*, Boys' League Prexy."	8
D. Class officer, chairman of a temporary committee.	"The committee chairmen for the senior luncheon are as follows: *Marie Anderson*, decorations, Ruth Roberts, publicity. . . ."	6
E. Member of a minor or temporary committee, also mentioned.	"The clean-up committee this term included *Art Dawson*, Roscoe Benson. . . ."	3
II. *Athletics*		
A. Athletic star, state-wide recognition.	"*Sid Murphy*, 6 feet 5 inches, all-around athlete, receives scholarship to Harmon University."	10
B. Special contribution.	"*Bill Hewitt*, veteran of last year's team, shows extreme promise on this year's gridiron. . . ."	7
C. Team member.	"Tomorrow's line-up against Polytechnic will probably be . . ."	
III. *Intellectual Leadership*		
A. Editor of the daily paper.	"*John King*, L-12, has been chosen editor of the 'Daily U-N-I' for the fall term."	6
B. Important signed contribution, editor of other minor school publications, staff position.	"East Bay writer picks San Carlos School to win grid race. *Rod Brown*, sports writer for paper, picks winning teams. . . ."	5
C. Member of group contributing.	"A protest against middy regulations . . . signed, *Lillian Marvis*, *Beatrice Vanetti*, etc."	4
IV. *Arts, Drama, Music, Dance*		
A. Outstanding talent displayed.	"*Glen Marks* wins summer scholarship to Franklin College of Arts and Crafts."	5

TABLE 1 (*continued*)

Category	Example	Assigned weight
B. Special contribution, performance.	"*Ann Mason, Clair Wilson* and *Gladys Carson* won first place on Bennie Walker's Amateur Hour of April 18."	4
C. Also played, accompanied group among those trying out.	"A group of students from the voice class including *Katherine Boyd* . . . will sing on tomorrow's assembly program."	3
V. *Interest Clubs*		
A. President.	"*Wes Rogers,* president of the Pet Club announced that a puppy would be raffled off at the School Fair next Friday.	2
B. Member.	"The French Club attended the Campus Theater to see "Dark Eyes" last Saturday. Among those who attended were *Lillian Mansfield* and . . ."	1
VI. *[illegible]llaneous*		
A. Extraordinary, unusual, outstanding event.	"'And the King of the Carnival is none other than our own *Karl Envick.*"	10
B. Special mention of interest.	"*Bernard Thomas* again displays his private gun collection. . . ."	7
C. Member of group mentioned.	"Credit goes to the following students who have been associated with office mimeographing, *Marion Vallejo,* Charlotte Smith. . . .	2

each sex who were in the top 20 per cent of the distribution of weighted scores, it was found that all but one of each sex would also have been chosen if the raw frequencies had been used.

The resulting selection includes not only the student body leaders, but a wide variety of persons in other social rôles who used their talents or energies in behalf of their high school associates.[4]

Indeed, followers, the so-called "middle" leaders, or "second level" as well as top leaders, are represented in this group since, as the discussion will show, some students by dint of zealously carrying out the committee work prescribed by an executive officer (chairman, editor, president, etc.) were frequently recognized in this middle leader or follower capacity. There were also those who caught the public attention because of their special talents or energetic services which were not necessarily associated with leadership ability. In the adult community, persons who espouse causes, get out the vote, start fads, or win decisions, vary greatly in talents though all may frequently get their names in the paper.

---

[4] "The high school with its athletics, clubs, sororities and fraternities, dances and parties, and other 'extracurricular activities' is a fairly complete social cosmos in itself" (11, p. 211).

Differences in purposes and abilities which led to greater or less success in their participating or non-participating rôles will be discussed in the presentation of the records of individual students. But group comparisons will be considered first. The question was raised earlier: What are the major differences between those students who manifestly identified with the purposes of the school to promote peer group interaction through extracurricular activities, and those who were unresponsive to this socializing influence? The hypotheses relevant to this question are implicit in the following selection of data categories:

Table 2 lists the variables used in comparing the members of the two extreme groups.

C. Results

1. *Skeletal Age*

Chronological age was not a differentiating factor. Skeletal age, however, was relatively advanced among the high-mention boys and among the

TABLE 2
VARIABLES USED FOR GROUP COMPARISONS

1. Chronological Age.
2. Skeletal Age.
3. Total strength scores (right hand grip, left hand grip, thrust and pull with dynamometer) (8).
4. Mental Age and *IQ*.
5. Adjustment Inventory Scores (U. C. Inventory I) Total score, and subscores for Family, Social and School Adjustment, Fears, Inferiority Feelings, Physical Symptoms, Generalized Tensions (16).
6. U. C. Interest Test Scores. (Athletic, Social, Mechanical-Scientific, Intellectual Interests) (9).
7. Attitudes Tests (Selected Items) (1).
8. Reputation Measures: U. C. Reputation Test. Ratings by classmates of appearance, leadership, bossiness, talkativeness, restlessness, attention getting, daring, sense of humor, popularity, enthusiasm, friendliness, happiness, activity (15).
9. Ratings of behavior in social situations by staff members (13).
10. Ratings on home and neighborhood. (Census Tract Data and Home Ratings.)
11. Drive Ratings: Autonomy, Social Ties, Achievement, Recognition, Aggression, Succorance, Control, Abasement, Escape (5).

low-mention girls (Table 3). For the total group of 56 boys of the Adolescent Study attending this school, a correlation[5] of .42 was found between skeletal age over *CA* at 15 years, and rank order of mention, with a significance level of .005. For 52 girls the corresponding correlation was —.42 (significance level .005).

TABLE 3
SKELETAL AGE AT 15 YEARS

	Boys		Girls	
	N	Mean	*N*	Mean
High-mention group	12	16.43	12	15.43
Low-mention group	9	15.46	9	15.97
Significance of difference		.02		.10

Other reports from the Adolescent Study indicate that in this middle period of adolescence early-maturing tended to be a factor of advantage for the boys in our group and a disadvantage for the girls (6, 10). This is to be explained partly on the basis of the timing of individual development in relation to the group norms. Girls who in general are early-maturing in relation to boys, as well as to other girls, find it a disadvantage to be at the extreme in this respect. They are conspicuously large in relation to other girls in their classroom and decidedly so in relation to boys of their own age. For adolescent girls in our culture this is not an admired characteristic. Boys, on the other hand, find that early maturity puts them on a par with the majority of the girls in their peer group in size and in the accompanying changes of interests.

The relationship between maturity level and recognition status in high school is shown in the accompanying four figures. Figure 1 illustrates the findings in regard to boys. Although only 20 per cent of our total population of 90 boys were accelerated more than one year in skeletal age, they con-

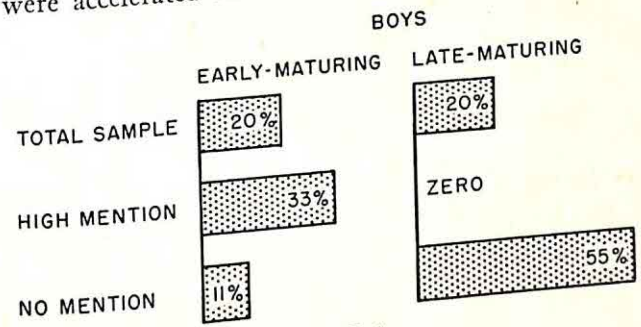

FIGURE 1

---

[5] Pearson Product-Moment.

stitute 33 per cent of those having great "newsworthiness" (10). None of the late-maturing boys appeared in this prestige hierarchy. Among the no-mentions, five times as many were late- as early-maturing (55 vs. 11 per cent).

Figure 2 shows corresponding data with regard to girls. A reversal of prestige position, in relation to maturity level, is clearly shown. Early-maturing girls fall more frequently in the no-mention group, while the late-maturing dominate the high mentions to the extent of about twice their expected quota.

Figures 3 and 4 provide cross-comparisons for boys and girls.

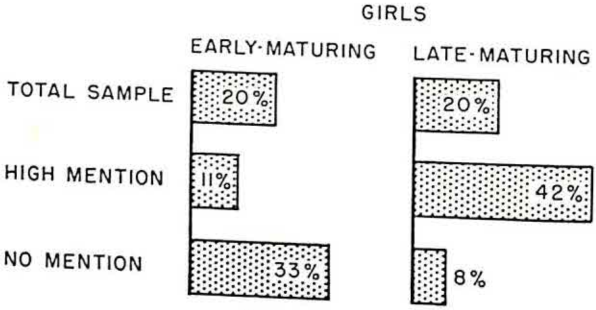

FIGURE 2

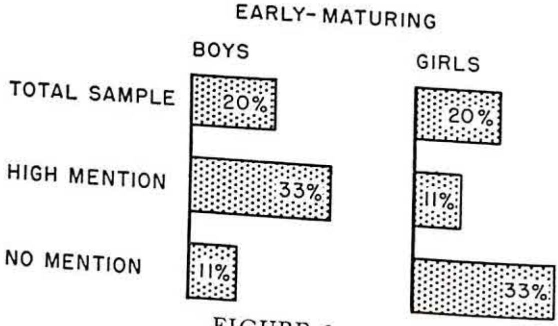

FIGURE 3

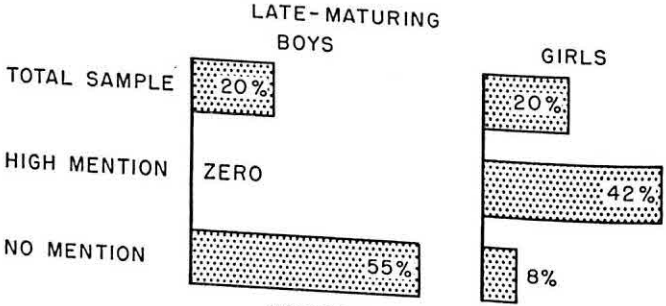

FIGURE 4

2. *Behavior Ratings*

It is not surprising that the two contrasted groups of this study were differentiated most clearly by estimations of their social behavior characteristics as displayed in group situations. It was the recognition of social behavior in connection with school activities which determined their newsprint value. Those who came frequently to the attention of classmates through prominence in extracurricular activities and those who escaped such notice, practiced behavior patterns and social rôles which distinguished between them in other situations as well.

With no knowledge of information based on the school paper, staff members who observed these adolescents on playgrounds, in school halls, club meetings, informal and formal social gatherings, school excursions and the like, ascribed certain traits to members of one of the extreme groups significantly more often than to the other (13). Both boys and girls in the prominently mentioned group were more frequently rated by adults as making a good impression on the basis of appearance and grooming (differences significant at the .02 level). Other descriptions based on adult ratings differentiated the girls' groups significantly (at the .05 level) in regard to their expressive behavior.[6] The frequently mentioned girls were considered to be more talkative, active, peppy, animated, and eager. Boys showed a tendency in the same direction but the difference was not significant. Other reports from the Adolescent Study have suggested that expressiveness, as defined here, is socially approved to a greater extent for girls than it is for boys (6, p. 78).

Girls and boys in the frequent-mention group were rated as buoyant (relaxed, cheerful), poised (assured, confident), and as having prestige (others seek his approval, expect him to take the lead; has a marked effect on the group). Table 4 and Figure 5 present these findings. The data are in terms of standard scores, in which 50 represents the average for the total group, and the *SD* is 10. Although boys and girls show differences in the same direction on each characteristic, the high- and low-mention groups are more clearly differentiated among the girls.

---

[6] The description of expressiveness is as follows: Very self-expressive; overtly active practically all of the time including verbalization, gross movement, aggressive contact with physical environment; eager, animated facial expression and bodily movements. The other end of the scale was described in these words: Extremely unexpressive, very little overt movement; silent; stationary, indifferent attitude; idle; stolid, listless expression. Newman, F. B. (13).

TABLE 4
OBSERVERS' RATINGS: STANDARD SCORES: MEANS

	Boys			Girls		
	High	Low	Sig. diff.*	High	Low	Sig. diff.*
Expressiveness	52	47		58	43	.05
Buoyancy	53	45		59	43	.05
Poise	58	48	.10	67	43	.001
Prestige	54	46		60	40	.01
Appearance	60	45	.02	64	45	.02

*Wilcoxon Rank Test.

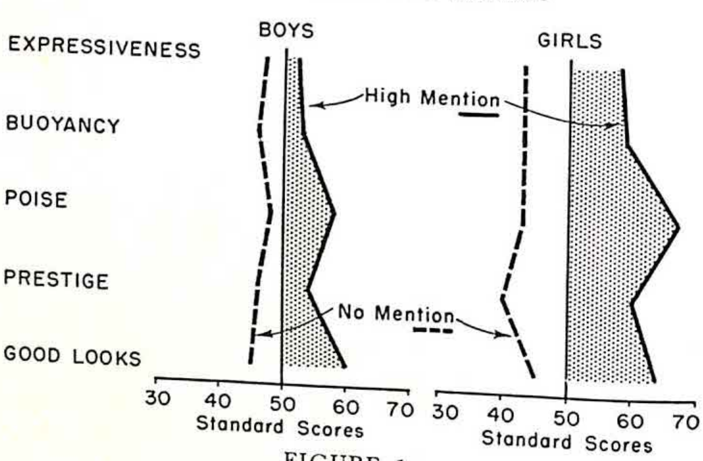

FIGURE 5

3. *Reputation Measures*

It may of course be expected that students who are centers of influence in extracurricular activities will have reputations with classmates, as well as with adults, quite different from those acquired by the non-participating. This was revealed by the Reputation Test, a "Guess Who" rating technique administered in the classroom (15). Those whose names were entered in the school paper most frequently were also judged to be well above the average in all prestige traits. The frequently-mentioned boys, for example, averaged approximately 2 *SD* above the group mean in class estimates of popularity ("here is someone that everybody likes, others are glad to have him around"); friendliness ("very friendly, has a lot of friends, is nice to everybody"), and humor. Among boys, the no-mention group tended to be overlooked in classmate nominations, rather than to have explicitly unfavorable reputations.

As will be shown in the later discussion of individual cases, lack of mention in the school paper did not serve to select a group considered by their classmates as "queer" or as having markedly distasteful qualities, although negative scores were relatively frequent in the case of the no-mention girls. Table 5 and Figures 6 and 7 present the data for reputation traits.

TABLE 5
REPUTATION MEASURES: STANDARD SCORES: MEANS

	Boys			Girls		
	High	Low	Sig. diff*	High	Low	Sig. diff.
Popular	69.20	50.62	.02	69.83	53.59	.001
Friendly	68.83	44.63	.02	72.25	46.67	.02
Leader	65.00	48.00	.02	58.17	33.56	.02
Daring	65.00	49.00	.001	61.33	42.33	
Enthusiastic	59.83	54.00		69.16	52.33	.10
Talkative	50.16	47.16		56.83	39.67	<.02
Humor	71.00	53.25	<.02	54.00	36.00	<.10
Good looking	64.25	50.88		65.75	44.11	.02

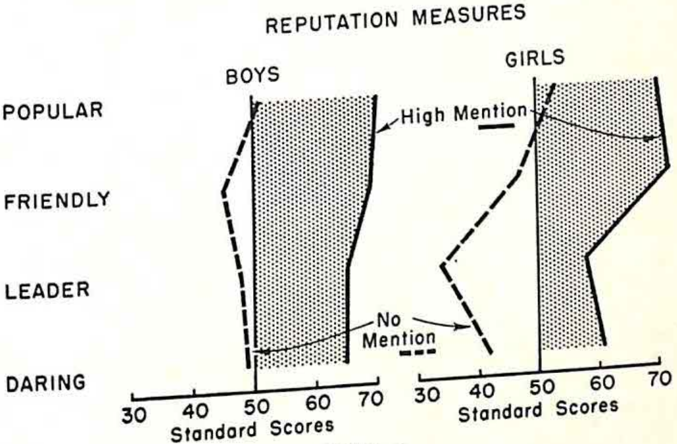

FIGURE 6

4. *Drive Ratings*

Toward the end of their high school careers, all of the study members were rated on a series of motivational traits based on Murray's needs. The ratings were made, independently, by staff members most competent for this assignment because of their intimate association with the group over a period of 4 to 7 years.[7]

Both among boys and girls, members of the high-mention group were

---

[7] The nine drives are presented in Table 2, Item 11.

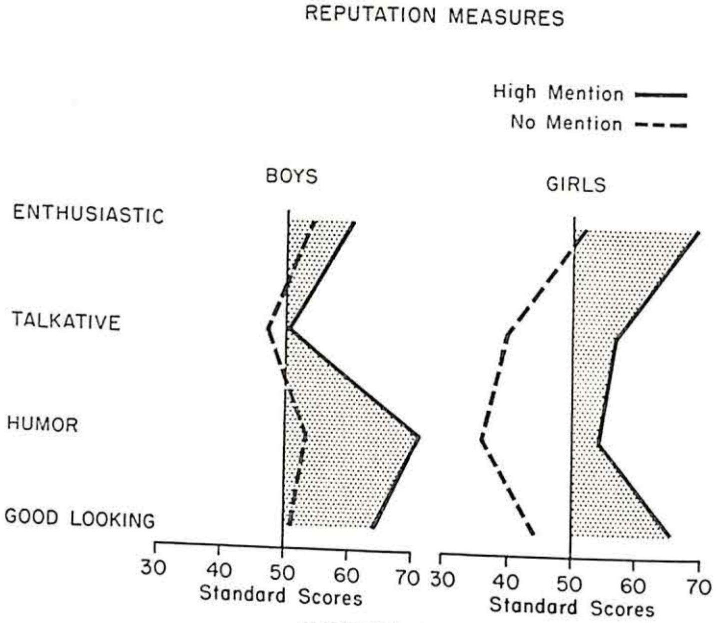

FIGURE 7

judged to have relatively strong drives (a low rating on the 5-point scale indicates a strong drive). The differences were significant, however, only among boys in the need or desire for recognition (drive to excite praise and commendation) and for control (to control one's environment, by suggestion, persuasion or command). The comparative data are shown in Table 6 and Figure 8.

TABLE 6
DRIVE RATINGS: 5-POINT SCALE: MEANS

	Boys			Girls		
	High	Low	Sig. diff.	High	Low	Sig. diff.
Achievement	2.3	2.9		2.5	2.8	
Recognition	1.9	3.2	.02	2.1	3.0	.10
Control	2.6	3.8	.01	2.3	3.1	.10
Aggression	2.8	3.1		3.0	3.2	
Social ties	2.4	3.1		2.2	2.6	

5. *Intelligence*

The foregoing has dealt with the areas in which the most striking differences in the two groups of subjects were found. Of considerable interest, however, are the findings concerning measures on which no differences occur. Differences in intelligence test scores were not significant. For the total

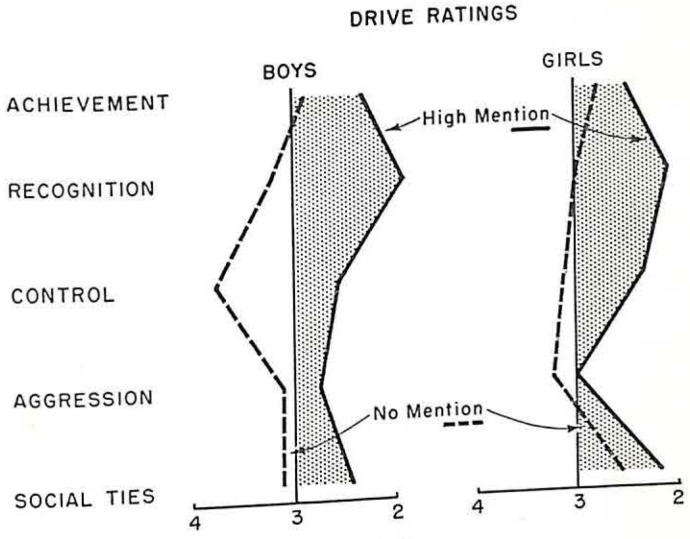

FIGURE 8

group of 57 boys and 54 girls, the correlation between *IQ* and number of newspaper mentions was .22 and .25 respectively (not significant at the .05 level).

A review of the kind of ego rewards which individuals achieved from extensive social participation suggests that the use of intellectual ability was only one among many possibilities. Several athletes, a cartoonist, a band-leader, a contralto singer, were among those who contributed, conspicuously, to the richness of school life though their intelligence scores were no greater than average.

On the other hand, some students with superior mental endowment were motivated to use this asset in more task-oriented and less social areas. There were four girls and four boys in this sample with *IQ*'s above 120 who were never mentioned in the "Daily." Here were potential resource persons who did not use their capabilities in the interest of high school extracurricular participation and leadership. Three of the four girls in this high *IQ* category who were non-participants were early-maturing. We have suggested that these girls tend not to fit into the school peer culture readily and it is not surprising that they failed to join in extracurricular activities. Three of the four boys in this same low prestige classification were late-maturing, a factor which has also been reported to be a disadvantage to social acceptance in adolescence. Maturity level is one of the more obvious in a matrix of complicated factors which we would need to understand in order to explain why these intellectually able individuals did not yield to the pressures to conform to the preferred socialization pattern.

6. *Physical Abilities*

Since in this particular peer subculture athletic skill in girls had little status value, we would not expect and did not find any relationship between physical ability and frequency of mention for the girls' groups. Among boys, a previous study of this sample of adolescents has shown that the muscular build which goes with strength and athletic ability is associated with classmate and adult ratings of "good looks" (10). Though "good-looking" was descriptive of those with high mention, the relationship found between strength indices (see Table 2 Item 3) and prominence in the school activities program was not significantly discriminating for boys. Several boys in the high-mention group were well-known athletes, but others (such as the sports writer on the school paper) had poor physical ability and attained recognition through other channels. Several boys in the no-mention group were relatively high in their strength indices, but never chose to join a team or to exhibit their physical abilities competitively.

7. *Socio-Economic Status*

Social status was found to exert only a small influence upon high school recognition, in terms of the criterion used in this study. There were to be sure certain prestige positions which tended to be filled by students of high social status.[8] However, these were not the top ranking offices, and participation in school activities attracted students from all social classes. Correlations of .16 for boys (58 cases) and .17 for girls (54 cases) were found between socio-economic status and frequency of mention. As in the comparisons of extreme groups, these were not significant at the .05 level. Perhaps the implication to be drawn is that the "common man" class of students tend to use the school as an educational medium for practicing leadership and participation rôles (upward mobile activities). Those of a higher social stratum are learning these rôles to some extent in the greater variety of social settings (outside school) which are available to them. Although prominence in school affairs gave a coveted mark of distinction regardless of social status, it is also possible that other prestige groups with which the students of higher socio-economic level identified themselves tended to divide their loyalties, so that less effort was expended in student body activities. Perhaps typical of these students is one boy whose attitude is illustrative:

---

[8] Measures used were ratings of home and neighborhood, census tract data, and use of community welfare agencies. At the time of this report (the depression era), the families of the children who were most active in school affairs used public service agencies as frequently as those who did not participate.

Reese, from an upper-class home, reported enthusiastically to his mother during his first week in High School: "There are four ways to make a name for yourself at Uni: One is in student body activities, another is in scholarship, a third is in the social life, and fourth in athletics. I have a good chance to get out in front in all of them." After a year during which time he had joined an out-of-school social club drawing members from several high schools, this same boy remarked: "If you belong to a good club it doesn't matter much which high school you go to—you meet the best guys from all of the schools in your club and would feel at home in any school." This boy's school loyalty had been supplanted by loyalty to his social club. Success in school and in extracurricular activities were weaker goals and ceased to command his best efforts.

In the geographic area of this study, schools and community agencies which offered resources for group participation were increasingly aware of the competition of invitational social clubs. Their problem was to offer equally satisfying social opportunities and goals, and to help the young people who came under their jurisdiction to coördinate and integrate their loyalties toward the various groups in which they functioned.

8. *Attitudes*

As has been indicated, members of the prominently mentioned group were frequently rated by classmates on the reputation test as outstandingly friendly. This seems to corroborate the impression of an over-all democratic atmosphere in the school and "a chance for anyone to become president." This should not imply, however, that the awareness of social distances did not exist. The following note by a staff member illustrates how some students felt about their reception into Senior High School. "*Jack* (a member of the frequently mentioned group) now in the 11th grade is less of a leader than he was in Junior High and he resents this. He bears a definite grudge against 'the hill students' and especially against the girls whom he describes as 'proud and snobbish.' " (Later he succeeded in marrying one of them.) Our attitude tests did not indicate significant differences between the groups in prejudice toward outgroups or in ethnocentric tendency (1).

9. *Self-Concepts*

Did these students have self-concepts which differed for the two groups? Data available for quantitative analysis of self-estimates were the self-report inventory, an interest check list, and attitudes measures (see Table 2). None of these yielded significant differences between the two groups. There was a slight tendency for girls who did not participate to report family problems and for boys who did not function in school events to mention more physical symptoms and to express greater dissatisfaction with school. Girls who did engage in school activities frequently tended to indicate more interests of an intellectual nature than did those who were isolated from the school extracurricular program.

Although there is a tendency to ascribe favorable qualities to leaders and others in prominent positions, being successful in these rôles is not primarily a matter of being well-adjusted, happy, or self-accepting. Nor did the successful students so reveal themselves. On the one hand, those who endorsed the socialization pattern would, presumably, experience ego satisfaction in achieving their goals. On the other hand, tension-producing conflicts would be involved for some, in learning the behavior expected of them in their selected rôles. Adults who rated these boys and girls did not describe the high-mention group as particularly social or good-natured. Nor did they attribute to them the kind of drives which are associated with a need for warm social ties or self-reliance. In describing their goal-seeking behavior, the staff differentiated between achievement needs "to increase self-regard by successful exercise of talent and ability" and a desire for recognition—"to excite praise, command respect, prestige, social approval." The participating group were considered to exhibit the latter in a greater degree (see Chart 8). Individuals who rated high on the drive for achievement may have tended more toward "inner-directedness," while those with high scores on need for recognition may be thought of as "other-directed," depending upon the approval of peers to define success for them (14).

10. *Rôle Patterns: Individual Descriptions*

As has been indicated earlier, newspaper publicity accrues to or is sought by persons who provide a wide variety of community functions and portray numerous social rôles. For this study three types of rôles have been roughly designated as among those shown by the most frequently mentioned group. The individuals who function in these rôles may appropriately be described as "the leaders," "the performers" and "the willing workers." An individual may combine these rôles in greater or less degree and with varying success.

Leaders were those who won elective offices and carried out the tasks of student body government. There were three boys and one girl in our sample who qualified in this category.

Karl, student body president, represents those who have rather generalized leadership ability and have accepted the prevailing cultural goals. These individuals usually hold student body offices and count their followers in large numbers. They have accepted the dictate "to amount to something" and "to be somebody." Karl was an only child from a lower middle class, Protestant, first generation Northern European family. Parental goals for their son were those typically ascribed to adults who seek acceptance and improvement of status in the middle class culture: assumption of responsibility, conformity to accepted social codes, and individual achievement. With other outstanding leaders among Karl's peers, such factors as the stress of family position or the example of older siblings may have weighed more heavily in furnishing motivation for performing leadership rôles. For Karl, his family expectations were to be accomplished through the prescribed channels of home, school, church, youth groups. His mental and physical endowments (he was "college material," well-built, and early-maturing) gave promise of his being able to achieve these goals. It was said that Karl's "impressive physical development did not allow him to relax into easy mediocrity" (he was the tallest boy in the study). At home there was a tendency to protect Karl because of his early growth spurt and conspicuous physique. He did not enter serious athletic competition until the late high school years. But at school, from the earliest grades he was selected by teachers and principals for leadership rôles. Before the end of his elementary school years, he was recognized as a leader among his peers.

In general, Karl's personal-social characteristics exemplify those which distinguished the student body leaders as a group, described by teachers and staff members in such terms as these: "not obviously dominating but directing," . . . "officiating in a paternal way," . . . "carrying the load of the conversation with adults."

More than the average student, Karl and his kind are willing to make approaches to young people outside their accustomed circles; to do the unpopular chore. "Voluntarily assumes positions in the game which no one else wants," was noted of one boy in this group. Of others it was noted: "Treats the boys tolerantly" . . . "speaks cordially to everyone as she goes down the hall" . . . "went to the physical exam with 'Dutch'" (an isolate) . . . "spoke with the girls who were being ignored by her friends."

In spite of the appearance of buoyancy, none of these student body leaders escaped a certain appearance of striving. Occasionally the notation "worried" appears. More frequently such phrases as the following are used: "Takes responsibility seriously" . . . "definitely strives for leadership position" . . . "popular as a leader but has to work for prestige."

On the self-report inventory, these young people did not describe themselves in such a way as to indicate high self-acceptance. Admission of fears, physical symptoms, tensions, and feelings of inferiority suggest that at least some of the striving for status is motivated by compensatory feelings. It also implies the ability to internalize emotional behavior and to tolerate anxiety (7).[9] We might catalogue these leaders as standard bearers; they had assumed responsibility for carrying on—mid snow and ice if necessary—the standards which their high school peers recognize and accept as those approved by the adult culture as well as their own peer subculture.

A second group of students receiving high mention in the school paper are "the performers." They rely upon "talent" to reach their goals. The athletes would come first in this category but it also included the musicians, cartoonists, dramatists, and writers. The personal approach played a less important rôle in their functioning and they actually varied widely in their social appeal. An athlete hero who also had social facility might be elected to leadership positions. One such student, *Bill*, was elected class president.[10] On the other hand among the "performers" was one boy who received a very low score for "friendliness" on classmate ratings. He won his laurels by his clever, though often biting, cartoons. One of the girls with "talent" was described by adult observers: "Paula has a somewhat snobbish manner which does not attract friends but does suggest an air of assurance, which is effective in some social situations."

A few in this talent group had aimed at the top social-prestige hierarchy (primarily represented by membership in invitational social clubs) and with reluctance had lowered their sights to suit their talents and social status. "Nancy (successful in dramatic rôles) had a great deal to say about how discouraging it was to have the 'hill bunch' run the school." Among these students, as with the standard bearers, there is evidence of anxiety to attain recognition: "There has always seemed to be some underlying irritability that keeps Jane in action." . . . "Tony seems ridden by his ambition to be a popular band leader." . . . "John was apologetic about not having enough time to work with his committee."

Tony, the band leader, represents the group who won recognition through their entertainment value. While Karl, student body president, was mentioned in every category of school activities except that of "The Arts," Tony participated exclusively in the area of his talent. He received publicity only in relation to his band. Some of the forces at work in motivating Tony are suggested from data in our files. Tony was among those students in whom the desire to be well liked was positively reflected in his behavior and furnished one of the strong forces urging him toward his goal. Another emotional impetus was his love of music, a symbol of his father-identification. A third may have been what he meant by "ambition," a physiologically high irritability which at times was drained off in nervous mannerisms, at other times consumed his energy, and threatened his health.

---

[9] It will be recalled that no significant differences were found between the two groups in adjustment scores.

[10] Karl became an outstanding athlete also in his senior year, which contributed to his prestige. However, it was not the basis on which his status was primarily built.

"To be a band leader," Tony said, "you have to have personality—to be able to talk to people and sell your band, unless you are a terrifically outstanding musician." Was Tony saying that he had to sell himself because he was not "terrific"? "And," said he, "you have to have ambition—you can't just have musicianship and personality; you have to *work*."

Prosaically a fourth factor which helped to mold Tony's choice of expressive medium was circumstance. When opportunity knocked—Tony was ready. The factors associated with Tony's choice of activity may best be reported beginning with the counselor's summary of Tony at the end of junior high school:

> Tony belongs to an ambitious and cultured Greek-American family. His father is a musician. For the first two years in Junior High School he was a quiet, tense, conscientious boy anxious to do what he could to get along well in school and with the peer group. He was accepted but not particularly popular. He was so anxious to be well-liked and to please everybody that he was really too submissive, both to adults and to his own classmates. He was asked to fill in at a few parties when he was in the H8 and at the beginning of the L9, but he was among the least popular as far as the girls were concerned. Then the orchestra was started, sponsored by the study staff.
>
> Tony became the center of and the leader of seven boys in this group. Everyone liked him as soon as he stopped being too anxious to please and acted more like the other boys. The dance orchestra became *his* orchestra. Bill began to pal around with him. It was when Bill was at the height of his popularity with both boys and girls. Tony rode to popularity with him, as Bill had certainly selected him for his best friend. Tony became less quiet, not so overly conscientious and anxious to please, more natural—though still a bit excitable and tense with his newly found musical and social life. With the opening of the H9 semester he had climbed to a position among the most popular boys, both among other boys and with the opposite sex. The girls who like him now, like him, I think, for his prestige as a dance orchestra leader. Several are honoring him with their pursuit. But, as he puts it, he doesn't want to get in love again. (An earlier episode claimed his emotions and was unrequited.) He is very much absorbed in music,

> learning the trombone which his family, at a great sacrifice, managed to buy for him.
>
> In the 9th grade Tony started the college preparatory course and will barely finish his Latin. Algebra he finds extremely difficult. This year's work has made him decide not to count on going to college and law school. He plans now to take a business course and keep on with his music. This change does not bother him as it would have a year ago when he was straining to get ahead academically and socially.

Thus Tony was well on his way socially when he entered high school, and became one of those frequently mentioned in the "Daily." Tony's need for social ties was evident, but others in this group with "talent" were not similarly motivated. *Walt* who wrote for the school paper and had the highest *IQ* in the group had a turn in his social fortunes which was practically the reverse of Tony's. From a grade school youngster with satisfactory friendships, Walt became more and more of a problem to himself in the adolescent years and by his grouchy, sullen, and arrogant behavior alienated his former good friends. His mother commented, "He's a know-it-all, and he argues everlastingly." The psychologist who administered the Stanford-Binet test wrote: "Walt interpreted the second proverb: 'You must not cast pearls before swine' as 'Don't waste your talents on things that don't deserve it.' This epitomizes his attitude. He did not want to appear to be taking the trouble to think out correct answers, to be the least bit concerned about the results, or to give the examiner the satisfaction of having him consider the task of any importance."

Walt's week-end diary written in the 12th grade described work at the newspaper office, going to a "show" alone, developing and printing pictures, alone. Although unhappy and socially isolated from his age group, Walt presumably gained satisfaction from the recognition gained through the exercise of his talent. And his classmates accorded him a measure of honor and fame on the basis of his ability and in spite of his condescending and critical attitudes.

Still a third category proved to be useful in classifying some of the high-prestige individuals: The "willing workers" or "middle leaders" . . . those students who are committee members *ad infinitum*—or who occasionally head small subcommittees where they are delegated to carry out a part of a larger project (such as a function at the Class Day Activities) but are practically never elected to major appointments. Such a girl was *Ruth*, a late-maturing, middle class girl of superior ability, who wrote in her 12th grade autobiography: "I had always admired certain girls who seemed to know how to run things at school or who had fun helping on committees. It was not until the 11th grade that I had my first chance and served on the poster committee." From this time on until graduation, Ruth was an indefatigable committee member.

The "willing workers" were less colorful as well as less stimulating than their more dominant classmates. Nevertheless they were imbued with the importance of participating in the high school social life. Wrote one of these:

> A girl in high school should enter school activities of all kinds in order to have a well rounded social pattern. She should, if possible, be a leader, be on committees and join organizations. Popularity plays its part in a modern girl's life. We all like to be popular and the best way is by getting in and doing things to make ourselves known. People come to know us by the activities we participate in and the clubs we belong to, that's why it's a good idea to enter all social affairs.

By dint of efficient or energetic participation these "workers" created a demand for their services, and were frequently appointed to committees; a fact duly reported in the school newspaper.

This high school was exceptional in the opportunities available to young people to practice leadership and engage in peer group social activities. It was stated earlier that nearly 90 per cent of the student body engaged in some school-connected activity. What then of the adjustment of the young people whose share in the school life was so meager as to bring no recognition?

For several who were in a "negative phase," failure to accept the high school pattern was merely an extension of a generalized rejective attitude. This is illustrated for example by *Florence* who was a "sturm und drang" adolescent, and a glum, sober, self-imposed isolate. An observer recorded: "She dislikes the University visit (part of the study program) because it imposes social adjustments which she is unable to make. . . . Today she had to go to the physical examination with Amy, a bad arrangement since Amy, though not as able a student as Florence, is poised, gracious, friendly—qualities which Florence would love to have but at the same time despises herself for wanting." In college Florence developed an absorbing interest in a professional career and married a man with similar ambitions. She is now a wife and mother as well as a professional partner with her husband. Her life seems rich and satisfying. But as a high school student she rejected the avenues of communication open to her through the school and peer group. At one time she remarked that her friendship with her older sister was more convenient than trying to make and keep friends outside the family. After high school she admitted that she had wanted to join clubs and go out for activities but blamed her sister for not taking the lead in this. Thus did she excuse her non-social existence. (Her self-report inventory revealed her tendency to project.) Florence is one of several in the study group who reacted to difficulties in their adolescent years by behaving negativistically toward others. Some individuals like Florence handled their adolescent adjustment problems by negativism or withdrawal. As adults they may now find satisfactions in professional achievement or in family living. Their social contacts are primarily with fellow workers or the family group.

Angela, another non-participating student, was a shy but bright, eager, responsive person. Observers noted: "Angela takes a docile rôle." . . . "She is unobtrusively interested in all that the others are doing or saying" . . . "she is pleased at small attentions" . . . "others seem to like Angela well enough but pay her little attention, and she seems to expect none" . . . "no one chose her for a partner" . . . "likes to be with others but does not seek recognition, nor has she enough stimulus value to be actively accepted or avoided." Angela was a motherless child, below par in health and physical vigor, tolerated rather than loved in a foster home. Like Angela, several others in the non-mention group were retiring. The heavy hand of repression had early placed its stamp upon them.

Although only five girls in the study have not married to date, Angela and two others of these are from this non-mentioned group.

Among the non-participating boys, *Eugene* was an early-maturing, tall, lanky, blond boy with a pale, rather homely, sad face. His health had not been good and he tended to complain of aches and pains. The youngest of a family of five and the only boy, his parents, older than the average of parents for high school boys, subscribed to the adage "Satan finds some mischief still for idle hands to do." The family had a financial struggle and Eugene had been deprived of money to spend on adequate clothes even when he earned the money himself. His parents kept Eugene busy with housework and carefully scrutinized his private and social life. Was it lack of enterprise or of security (characteristics of his high school days) which was suggested in this statement from his autobiography? "I am still living in the same house in which I was born and hope to keep on living in it as long as it stands."

In the 11th grade this comment was written by an adult: "For the last two years Eugene has not been mentioned by anyone as a best friend." The sense of isolation from a group of contemporaries is revealed in a composition written by him.

"The dog that shared six very enjoyable years with me died but a few weeks ago after being poisoned by some person still unknown. He was small but showed more affection than any human that I have ever known, or shall probably ever know." This was Eugene—who never got his name in the paper in the three years of his high school career.

Vincent started adolescence handicapped in physique, family background, and intellectual ability. He received few mentions either favorable or unfavorable on the reputation test. His early maturing, however, brought new strength and power to his physique, helped him to face his situation in junior high school and to accept his limitations. As a high school student now stable and more assured, he was described as "quiet, almost sedate . . . seems like an uncle to the other boys. . . . He joins their games but as an adult who tolerates play, rather than as a boy who craves it . . . reserved yet friendly . . . listened quietly to others . . . rarely spoke . . . relaxed and self-contained." His responses on a self-report inventory were consistently on the well-adjusted side. His vocational aim, accepted as practical at a relatively early age, was to become a truck driver like his father. He has realized this ambition and is now working for the same company as his father. Although observed to be very shy with girls in high school, he is now married and a family man. There is little evidence that lack of recognition while in school has had serious detrimental effects.

11. *Discussion*

What has happened to the active high-mention group 15 years out of high school? Some had four years in college. In this situation, similar to high school in its emphasis on campus activities, the pattern of participation continued. This was more true of the girls than of the boys. The latter, especially those in professional schools, were taking studies more seriously and so had less time for extracurricular activities. Seventy-five per cent of the high-mention group went to college as compared with 55 per cent of the no-mention group. (Forty-five per cent of the total study group went to college.)

The armed services provided another screening situation for the two groups. Sixty per cent of the frequently mentioned boys became officers, and only one member of the no-mention group. The only woman officer from the group of girls was in the high-mention category.

Fifteen years after high school, in a follow-up study, the boy who was remembered by his associates most frequently (from a high school photograph) was the band leader, Tony. Next was his best friend, Bill, the athlete-leader. Karl, the outstanding student-body leader was third. Some post-high school events probably influenced this order. Tony is still the leader of his own band which keeps him before the public. At one time, Bill received local mention in the papers for a war history involving injuries. But there was also a colorfulness about the first two which was somewhat lacking in Karl's appeal. Tony was warm, urgently in need of acceptance. Bill was the adolescent hero, in build, looks, masculinity, athletic ability—even in the possession of temperamental qualities which stamped him as human. Karl was admired, respected, trusted, perhaps not so well loved.

If a ranking of the individuals in the study group on the basis of prestige were to be made at the present time, there would be some decided shifts. One young man who can now properly be described as a celebrity was below average in number of mentions in high school.

The high-mention group furnishes more individuals who have achieved status in jobs or community, but in the no-mention group at least a third of the girls and a fourth of the boys have attained a degree of recognition comparable with any in the group who were so much more conspicuous in high school. The pattern of social interaction represented by participation in extracurricular high school activities provided many adolescents with social rôles which they found intrinsically congenial and likewise useful as a step toward other kinds of satisfying interpersonal relationships.

Some students who were not ready or able to practice this approved social behavior in high school felt insecure and unappreciated. Some found more personally appropriate avenues of communication through task-oriented activities, family, church, or community sponsored social activities or through small intimate friendship groups. In general, the patterns of social participation whether through leadership, entertainment, or service tended to persist into college and adult life. However, this varied with the opportunities available and the values placed upon such rôles in the individual adult's immediate environment.

At present a few young people are active in labor or church organizations. Among the married young women, leadership abilities are in some cases being used effectively by coöperative nursery schools.

All of the girls who were frequently mentioned in the high school paper are married and living with their husbands.[11] A few are now exhibiting more active public rôles and receiving more acclaim than they did as high school students. At present the more intimate interpersonal relationships of family life and the exigencies of a vocation are claiming the attention of these young adults who are now in their thirties.

---

[11] At this time 96 per cent of the girls and 90 per cent of the boys in this study were married. Nineteen per cent of the married girls and 6 per cent of the married boys have been divorced.

In general, opportunities to lead and influence of the kind available in high school are relatively lacking to these individuals as young adults. Perhaps we should put it another way and say that such opportunities are luxuries which they cannot afford. For the most part, neither their working nor their leisure hours are organized so that they have anything comparable with the number of interpersonal contacts which were customary in high school.

REFERENCES

1. ANASTASIOU, N. A Study of Prejudice in An Adolescent Group. Unpublished paper on file. Berkeley: Institute of Child Welfare, Univ. California, 1950. Pp. 22.
2. BAYLEY, N. Size and body build of adolescents in relation to rate of skeletal maturing. *Child Devel.*, 1943, **14**, 47-90.
3. BROWN, M. Leadership Among High School Pupils. New York: Columbia Univ. Press, 1933.
4. BROWN, M., *et al.* The University High School study of adolescents; basic philosophy. *Univ. High Sch. J.*, 1938, **17**, 67-69.
5. FRENKEL-BRUNSWIK, E. Motivation and behavior. *Genet. Psychol. Monog.*, 1942, **26**, 121-265.
6. JONES, H. E. Adolescence in our society. In *The Family in a Democratic Society*. New York: Columbia Univ. Press, 1949. (Pp. 70-84.)
7. ———. The study of patterns of emotional expression. *In* M. L. Reymert (*Ed.*), *Feelings and Emotions*. New York: McGraw-Hill, 1950. (Pp. 161-169.)
8. ———. Motor Performance and Growth. Berkeley: Univ. California Publ. Child Devel., 1949. Pp. 181.
9. JONES, M. C. The Interest Record. Berkeley: Univ. California Press, 1944. Pp. 63. (Mimeographed.)
10. JONES, M. C., & BAYLEY, N. Physical maturing among boys as related to behavior. *J. Educ. Psychol.*, 1950, **41**, 129-148.
11. LYND, R. S., & LYND, H. M. Middletown. New York: Harcourt-Brace, 1929.
12. MURPHY, G. The relationships of culture and personality. *In* S. S. Sargent and M. Smith (*Eds.*), *Culture and Personality*. New York: Viking Fund, 1949.
13. NEWMAN, F. B. The adolescent in social groups; studies in the observation of personality. *Appl. Psychol. Monog.*, 1946, **19**. Pp. 94.
14. RIESMAN, D., GLAZER, N., & DENNEY, R. The Lonely Crowd. New Haven: Yale Univ. Press, 1950.
15. TRYON, C. McC. Evaluation of adolescent personality by adolescents. *Monog. Soc. Res. Child Devel.*, 1939, **4**, No. 4. Pp. 83.
16. ———. U.C. Inventory 1, Social and Emotional Adjustment. Berkeley: Univ. California Press, 1939. Pp. 80.

Institute of Child Welfare
University of California
Berkeley 4, California

The Journal of Genetic Psychology, 1958, **93**, 113-117.

SOME PSYCHOLOGICAL FACTORS IN THE READING OF FICTION*

Departments of Psychology, University of Oklahoma and University of Wichita

Lee O. Thayer and N. H. Pronko

A. Introduction

Look at a page of fiction in an unfamiliar language. Nothing happens. Now look at a printed page of a particularly appealing "story" in English. What happens? We "see" the characters, we have feelings for them, or with them. We are happy or sad with the fortunes of the hero.

How does it happen that certain words and phrases, put together in a certain way, can elicit in a reader a "living picture" of the story being revealed by its author?

Furthermore, we don't *see* the same picture. It would seem that there must be contributing factors beyond the linguistic. We seem not to "*perceive*" identical cues from which to construct our images, and we seem to "*project*" into this image something of our own creation.

Since reading plays such an important rôle in our daily lives, an understanding of the processes involved is warranted.

B. The Present Study

This exploratory study was designed and conducted to provide data necessary to the design of future studies. For this reason the present study is meant to emphasize the relative significance of a wide variety of contributing factors rather than the rôle of any one factor. The final objective is to determine the fundamental psychological factors involved in the reading of prose fiction.

C. Procedure

One hundred twelve college students (31 female, 81 male) of varying ages and classifications, in beginning psychology and English courses, were asked to read five fiction excerpts, each of about one-half page in length. Then the students were asked to answer the following questions about the central character in each of the excerpts:

1. Did you have a good mental picture of ———?
2. Did you like ———? Why?

---

*Received in the Editorial Office on June 6, 1957.

3. Describe ——— in your own words, in any manner you wish.
4. Upon what details or ideas in the passage have you based your answer to number 3 above?
5. Why do you suppose you selected those "cues" which you did, and possibly overlooked others?

The students were then asked to check a personality profile chart for the central character in each excerpt. The profile consisted of 25 selected personality "traits": aggressiveness, anxiety, argumentativeness, bashfulness, business-like, cheerfulness, decisiveness, confused, coöperative, dullness, easy-going, egotistic, friendliness, hesitant, impulsiveness, introspective, laziness, meekness, quietness, resourcefulness, self-confidence, self-consciousness, shyness, talkativeness, energetic.

Fundamental to the investigation was the fact that none of the characters was described in any manner in any of the excerpts. None of the students had seen any of the excerpts before. The excerpts were unrelated, except that each dealt with some phase of a love situation, a general theme of interest to most college students. Each of the excerpts was chosen as being representative of a different narrative technique.

D. RESULTS

1. The reported clearness of mental images varied from the 59 per cent who saw the third character clearly to the 91 per cent who reported "seeing" the second character clearly.

2. Not all of the characters were equally "liked." The first character, for example, was "liked" by 80 per cent; the second character was "liked" by only 20 per cent of the students.

3. Character descriptions were free and unlimited by the design of the experiment. For each of the five characters, there was a community of responses for both physical and personality descriptions. Concurrence of opinion about the appearance and personality of each character was marked. Deviations were isolated.

4. The personality profile for each of the characters showed marked clustering of opinion, representing a high degree of concurrence or similarity of impressions.

E. INTERPRETATIONS

Certainly many determining factors are at play in producing an individual reader's reactions to his reading—social and cultural patterns, linguistic habits and preferences, reading abilities, "tastes," and so forth. Evidence of these factors was apparent in the responses to the fictional situations imposed in this study. The clearest evidence, however, isolated the *processes* rather than the content of imagining. What, then, seems to happen when we read fiction?

1. One approach to the results of this investigation suggests that the reading of fiction involves both perceptive and projective processes. Without the operation of *structural* factors in perception of these fictional characters, concurrence of opinion among students would have been at a minimum. Clustering of opinion in describing the characters would have been negligible.

Deviations from the general concurrence of eidetic impression cannot alone be accounted for by structural factors in perception. *Functional* factors—the present needs, wants, and feelings of the reader—in apperceiving these fictional characters are apparent. Though one of the characters may have appeared to a reader as young and pleasant (structural factors dependent upon past experience with the general type of fiction situation, the mood, tone, or language of the excerpt, and the similarity of that fictional character to other fictional characters), the same character may have appeared to another reader as young but unpleasant (functional factors, dependent upon the reader's present mood, feelings, wants, etc.). The second reader's apperception of that character as unpleasant is a deviation from the concurrence of the majority of readers that the same character is pleasant, and thus suggests the stronger influence of functional factors in perceiving the character. One of the most pertinent discussions of the definition and meaning of structural and functional perception is to be found in Hallowell (3).

Since no description of the characters appears in the excerpts, those details of description by the readers in which there is a significant concurrence may be considered to be dependent upon structural factors. In other words, those readers who concurred in their impressions were perceiving cues identified with common or similar past reading or other experiences. Those details of readers' descriptions in which there is a significant non-concurrence may be considered to be dependent upon functional factors. Those readers who did not concur in their impressions were perceiving cues which were identified with subjective feeling states or dissimilar personal experiences.

2. The results of this study also suggest that mental images in the reading of fiction are largely dependent upon the reader's projection of subjective interpretations into the situation.

If the fictional character is "liked," this seems to suggest, then, that character is endowed by the reader with the personality traits of what that reader would consider to be a "likable" person. If the fictional character is "disliked," then the reader seems to endow that character with what that reader would consider to be "undesirable" personality traits. If the reader neither "likes" nor "dislikes," then the mental image of the fictional character seems not to be clear to the reader, and the description provided by the author may thus be automatically accepted. Where an excerpt, by implication, intimates to the intelligent reader that the character is a "pleasant guy," then the reader, particularly if he neither "likes" nor "dislikes" the character, will be inclined to accept this implied description.

But if the reader "likes" or "dislikes" a fictional character, these results suggest, then the reader's image of that character will be relatively independent of the author's direct or implied descriptions of that character, and at the same time relatively more dependent upon the reader's projection into the fiction material what are to that reader the physical and personality traits most suitable to that reader's likes and dislikes in real people. In other words, if a reader really "hates" a fictional character, then that character is likely to appear to that reader with physical and personality traits considered "hate-able" by that reader, and this impression will be fairly impervious to the author's own description of the character.

Projection in the reading of fiction would seem to be, based upon the evidence of this study, dependent to some extent upon past reading experiences. It would seem probable that this is true because a certain number of reading experiences are required to establish the ability to identify and thus to discern those one likes from those one dislikes in fiction. A person who has read much and widely, for example, may be less likely to project a stereotyped pattern for a "villain" or a "hero" than will the person who has read little.

Another possible interpretation of the results discussed above is in terms of an interbehavioral theory of language as developed by Kantor (4). Whereas most of our responses, such as throwing a ball, are unistimulational, language reactions are bistimulational. Whether we talk or listen to someone's speech, we react to two things concurrently. When we ourselves speak, we are obviously reacting to the person listening to us and to the "thing" we wish to tell him about. When we listen, we react to the speaker's utterance and to the "thing" he is talking about (5).

Now, to come to our results. Since to a certain degree each individual's psychological history is unique and specific to him, to that degree will he react individualistically to the reading matter and its references. To the degree that our subjects had common backgrounds with the language and experiences "tapped" by the reading excerpts, to that degree did they show a concurrence of responses. This is an alternative way of explaining *both* the specificity and mutuality of the subjects' reactions in our experimental situation.

3. Students could not give answers to Questions 4 and 5. According to the one theory, this may indicate the inaccessibility of the perceptive and projective processes in the reading of fiction. Often, to support a descriptive detail, completely imaginary "cues" were "perceived" in the excerpt. Such behavior may reflect the complicated interplay of perceptive and projective processes in reading fiction.

Hallowell's study (2) suggested to him that the important element in the study of literature as a mirror of culture is the implicit rather than the manifest content. The same principle is true, it would seem, of the results of the present study. Manifest content of student responses to fictional situations is only a result of implicit "meanings" or reactions. For example, there is reason to believe, as pointed out particularly by Erickson and Lazarus (1), that some degree of perceptual defense may be operating in the structuralization of mental images from the printed page of fiction.

An interbehavioral explanation for the failure of subjects to isolate the linguistic variables that called out their responses would stress the relatively greater rôle of the reader's reactional biography and the incidental rôle of the reading matter. In other words, the experiences that the reader brings to his reading is more important than the black marks on white paper. Language can "trigger" a host of the individual's interrelated responses. In Emerson's terms, "He who would bring home the wealth of the Indies, must first take with him the wealth of the Indies."

We offer the above alternative approaches to the results of the present exploratory study of what goes on when we read fiction.

REFERENCES

1. ERICKSON, C. W., & LAZARUS, R. S. Perceptual defense and projective tests. *J. Abn. & Soc. Psychol.*, 1952, **47**, 302-308.
2. HALLOWELL, A. I. Myth, culture, and personality. *Amer. Anthropol.*, 1947, **49**, 544-556.
3. ———. Cultural factors in the structuralization of perception. *In* Rohrer, J. H., & Sherif, M., *Eds.*, *Social Psychology at the Crossroads*. New York: Harper, 1951.
4. KANTOR, J. R. A Survey of the Science of Psychology. Bloomington, Ind., Principia Press, 1933.
5. PRONKO, N. H. Language and psycholinguistics; a review. *Psychol. Bull.*, 1946, **43**, 189-239.

Department of Psychology
University of Wichita
Wichita 14, Kansas

Department of Psychology
University of Oklahoma
Norman, Oklahoma

The Journal of Genetic Psychology, 1958, **93**, 119-131.

THE RELATIONSHIP BETWEEN CHILDREN'S PLAY INTERESTS AND THEIR MENTAL ABILITY*

George Peabody College for Teachers

JAMES D. WANG[1]

A. INTRODUCTION

When one attacks this problem he finds that there are certain interpretive guides in earlier research to aid him. In fact several investigators have attacked this problem from one point of view or another. Among these studies we may note that Boynton and Ford (1) have made an analysis of the play activities of relatively bright and relatively dull children. From a time-card record kept of all activities, the following conclusions were drawn: (*a*) As regards the total amount of time spent in play, the average bright child was found to spend about 45 to 50 minutes a day more in play than does the average dull child; (*b*) there was no appreciable difference in the average amount of time the two groups spent in physical play; and (*c*) the average bright child spent more time in mental recreation than did the dull child.

Witty and Lehman (13) selected 50 children with an *IQ* of 140 or more for an intensive study. Each gifted child was paired with a mentally average child of like age, sex, and environment. The two groups were compared, and it was found that the gifted group and the control group of children demonstrated about the same versatility of interest in play and engaged in the approximately same number of play activities. The gifted children included in this study were found to be somewhat more solitary in their play than average children and tended to avoid certain types of vigorous play. Although the number of activities in which both groups took part was the same, the gifted group devoted much more time than the control group to reading.

Lehman and Witty (7) used the Lehman Play Quiz on groups of subnormal, normal, and relatively superior children and concluded that the bright children were less interested than the dull children in motor, religious, and social activities, the dull were less interested than the bright children in those activities which required reading and those which required a

---

*Received in the Editorial Office on June 13, 1957.

[1] The author wishes to express his gratitude to Dr. Paul L. Boynton, the President of the S. F. Austin State Teachers College, Nacogdoches, Texas, who has given insightful help in the preparation of this study.

sense of humor. Hollingworth (6) reported that the gifted children were more interested in games of intellectual skill than unselected children and that outdoor sports were almost as popular with them as reading.

Terman (10) found that the superior children did more collecting, and made larger collections of a scientific and historical nature, than did average children. The gifted children were found to possess more information about play, games, and amusements than the control group. He also indicated that the gifted showed greater interest in activities that required thinking than the control group and that they were slightly less social and more quiet. Dimock (4) analyzed certain data pertinent to the play pursuits of about 200 boys in Milwaukee and Kenosha, Wisconsin, and reported that boys who were superior in mental ability, who generally came from homes of higher socio-economic status, were more diversified in their play interests than the average boy.

Witty and Beaman (12) made a study which is focused in part on some of the inadequacies of the present curriculum, and suggest what they consider to be more appropriate content as well as justifiable objectives in guiding the play of the subnormal child. Two thousand, three hundred and six children, 804 girls and 1,502 boys, in 100 classes were involved in this study. It is of interest that boys outnumber the girls about two to one, and that the majority of the children are classified mentally as aments. It was concluded that many of the games frequently used in special classes did not fall within the limit of the pupils' abilities, nor did they represent the children's backgrounds in interest, and that the play activities in which subnormal children took part outside of school revealed less originality than did the extra-school activities of the normal child. Furthermore, the study indicated that both the girls and boys of special classes played and enjoyed outside of school those games which they were taught to play in school, and that the subnormal child did not report reading as an activity to which he turned frequently out of school, and seldom indeed did he mention this as a favorable school diversion.

Lewis (8) reported that gifted girls are more interested than gifted boys in reading novels, mysteries, in playing musical instruments, sewing and housework, dramatics, playing make-believe games, religious activities, and social clubs, and the gifted boys are more interested in reading science, history and biography, in active games and sports, and building things. McGehee (9) reported that retarded boys are more interested in active games, building things, driving cars, and working, and that retarded girls are more interested in reading novels, sewing, housework, playing make-believe games, and social clubs. By putting these two reports together, one may see that there are some activities which are in common for both gifted and retarded children, and that the main difference may be found in reading science for boys and in playing musical instruments for girls.

B. Method

The present analysis involves 1,800 children in the fourth, fifth, and sixth grades in certain Tennessee and Kentucky schools within a radius of 150 miles of Nashville, Tennessee. Both rural and urban cases are included. Within the grades there were 265 boys, and 253 girls in the fourth grade; 230 boys, and 296 girls in the fifth grade; and 367 boys, and 389 girls in the sixth grade.

The play interests of the groups were investigated by means of a play inventory of 108 items. The manner of constructing this list is comparable to that found in the better-known Lehman and Witty Play Quiz. The children's play interests are expressed in terms of likes and dislikes. The mental ability test which consists of only 20 questions was constructed in the same manner as certain of the well-known, standardized multiple-choice types. The scoring was made thoroughly objective. The most outstanding feature of this test is that it can be taken by the children in these grades within a total time of only six minutes. Since no more direct method of determining the validity of the test was evident, it was decided to compare the results obtained through its use with those obtained from administering the Henmon-Nelson Tests of Mental Ability, Form *A*, for Grades 3-8. The positive correlation between Henmon-Nelson and Wang test was .80 which seemed to be sufficiently high to justify the experimenter in using it in the present study. However, one further check was made on the test. A partial correlation was run between the two tests, eliminating or holding constant the effects of chronological age. This resulted in an r of .81. In light of these findings it appeared that the use of the short test was justified in the present study (11).

In order to test the relationship between the variables in question the highest 10 per cent, the lowest 10 per cent, and the middle 20 per cent of cases, as determined by the mental ability test, were sought for special study. Due to score distributions, exact percentages could not be obtained. It was believed that if a relationship should exist between the variables under consideration it was logical to assume that it should be thoroughly evident in comparisons of such extreme mentality groups as these.

C. Analysis and Discussion

1. *An Analysis of Play Activities Liked Best by Children of Different Mental Ability Levels*

It may be observed from Table 1 that the percentage values of play activities, 12, 32, 35, 37, 42, 65, 67, and 93 for girls, and 10, 24, 40, 42, 63, and 93 for boys, increase progressively with an increasing mental ability. Play activities 42, swimming, and 93, going to the movies, rather strikingly were the only ones of the 108 listed about which there was a progressively or consistently increasing expression of interest along with increasing intelligence, as here measured. Certain activities, as for example, playing musical instruments, reading, occurred so infrequently among the preferences of one sex group, in this case the boys, as to afford no meaningful basis of comparison.

It has been shown by the results of several studies along this line that children of higher mental ability are more interested in reading than are those of lower ability. Their interests in reading seem to be rather generally accepted. In the present analysis, however, it would appear that this interest is dependent upon the experiences, customs, and attitudes of the different sex groups as well as upon their intelligence. Thus girls of higher ability evince more interest in reading than those of lower ability, but boys, irrespective of intelligence, give little evidence of reading being among their most preferred recreational activities.

Contrary to what some might have anticipated, only three activities, 65, 67, and possibly 93, among the 14 activities which are preferred increasingly by increasingly intelligent groups, are of the quiet or somewhat sedative types of recreational pursuits. In other words, with increases in brightness there was an increasing interest in play of a rather active nature.

It may be seen from Table 1 that the play activities whose percentage values have a considerable or progressive decrease with an increase of mental ability are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 26, and 80 for girls, and 1, 6, 8, 41, and 59 for boys. Among these play activities, three are common to both the girls' and boys' groups. These are: 1, playing tag, 6, hide and seek, and 8, drop the handkerchief. It is very obvious that most of these activities require little intelligence in order to engage in them, and that many of them are lacking in lasting interest as compared with other play activities. In other terms, they appear to be rather immature forms of recreation. It is natural, then, that children of higher mental ability show less regard for activities of this type.

TABLE 1

A COMPARISON OF CERTAIN PLAY ACTIVITIES LIKED BEST BY AT LEAST 10 PER CENT OF THE CHILDREN IN ANY ONE OF THE THREE MENTAL ABILITY GROUPS

	Percentages of participation			Probable errors of percentages		
	Low	Mid.	High	Low	Mid.	High
(Girls)						
1. Play tag	18.6	9.8	6.8	2.6	1.5	1.5
2. Ring around the rosy	11.4	1.1	0.8	2.2	0.5	0.4
3. London Bridge	20.6	5.5	0.0	2.8	1.1	0.0
4. Farmer in the dell	25.8	6.0	3.4	3.0	1.2	1.1
5. In and out the window	22.7	6.5	4.2	2.9	1.2	1.2
6. Hide and seek	20.6	14.2	4.2	2.8	1.8	1.2
7. Hopscotch	11.4	6.0	5.9	2.2	1.2	1.4
8. Drop the handkerchief	17.6	3.8	1.7	2.6	0.9	0.7
9. Blindfold	14.4	8.7	3.4	2.4	1.4	1.1
10. Postoffice	7.2	13.6	8.5	1.8	1.7	1.7
12. Dodge ball	5.2	10.2	13.6	1.6	1.6	2.1
23. Play with dolls	15.5	11.0	11.8	2.5	1.6	2.0
26. Play jacks	16.5	14.2	10.2	2.6	1.8	1.9
32. Ride bicycle	40.2	53.0	55.1	3.4	2.5	3.1
33. Skate	31.0	38.1	34.0	3.2	2.4	3.0
35. Hike, or take long walks	6.2	13.1	14.4	1.7	1.7	2.2
37. Dance	15.5	23.5	27.1	2.5	2.2	2.8
42. Swim	11.4	20.8	34.0	2.2	2.0	3.0
43. Ride on horseback	21.6	38.7	34.0	2.8	2.4	3.0
48. Jump the rope	8.2	12.0	5.9	1.9	1.6	1.4
65. Play piano, violin, etc.	5.2	12.0	14.4	1.6	1.6	2.2
67. Read books and stories	1.0	2.7	11.8	0.7	0.7	2.0
93. Go to the movies	7.2	17.2	27.1	1.8	1.9	2.8
(Boys)						
1. Play tag	18.4	11.8	4.9	2.6	1.7	1.3
6. Hide and seek	18.4	16.0	7.8	2.6	2.0	1.8
8. Drop the handkerchief	14.6	4.7	1.0	2.4	1.1	0.7
10. Postoffice	3.9	11.2	22.3	1.1	1.7	2.8
15. Marbles	16.5	17.8	12.6	2.4	2.0	2.2
22. Play with pet dog or cat	8.7	10.1	6.8	1.8	1.6	1.6
24. Play with toy airplanes	4.9	10.7	13.6	1.3	1.6	2.3
29. Work with machinery	8.7	10.7	9.7	1.8	1.6	2.0
30. Fly kites	7.8	11.8	6.8	1.8	1.7	1.6
32. Ride bicycle	31.1	38.4	28.1	3.1	2.6	3.0
33. Skate	8.7	11.2	5.8	1.8	1.7	1.5
40. Fish	17.5	21.3	21.4	2.5	2.2	2.7
41. Hunt, or shoot just for fun	22.4	21.3	20.4	2.8	2.2	2.7
42. Swim	23.3	23.7	34.0	2.8	2.2	3.2
43. Ride horseback	26.2	30.8	26.2	2.9	2.4	3.0
56. Baseball	16.5	16.6	13.6	2.4	2.0	2.3
59. Boxing	14.6	13.0	11.6	2.4	1.8	2.1
62. Basketball	10.7	11.8	8.7	2.0	1.7	1.8
63. Football	31.1	42.5	54.2	3.1	2.6	3.3
93. Go to the movies	5.8	16.0	21.4	1.5	2.0	2.7
96. Watch baseball, football, etc.	5.8	4.1	10.7	1.5	1.1	2.0

When an analysis is made of the percentage incidence of preferences of different groups one finds some rather pertinent evidence of the way in which intelligence apparently affects certain of these choices. With respect to play tag, the decrease in interest is rather consistent among both girls' and boys' groups. With respect to ring around the rosy, however, although there is a progressive decline in interest, the most striking feature is not the lack of interest of the highest girls, but rather the extent of interest of lowest girls. In other words, there is relatively little difference in the preferences of middle and highest girls, but both of these groups differ appreciably from girls of lowest intelligence in their interest in ring around the rosy as a form of play. The same holds true with respect to the following activities: London Bridge, farmer in the dell, in and out the window, hopscotch, and drop the handkerchief, for girls, and drop the handerchief for boys. In fact, practically all of these would be eliminated from Table 1 but for the interest in them manifested by children of low intelligence.

When the 10 most popular activities were selected from each level of mental ability as contained in Table 1, it may be noted that play activities 32, riding bicycle, and 43, riding horseback, are among the best liked by all the children, both girls and boys in every mental level. Play activity 42, swimming, was favored by almost all the children except those girls of the lowest level of mental ability. These activities, then, are popular among all types of children, with little regard to the mental status of the child. Four play activities, 40, fishing, 41, hunting, or shooting just for fun, 56, baseball, and 63, football, which were among the best liked by boys at each mental ability level, did not appear in any designated mental ability group which was composed of girls. So in these we have activities which apparently are associated with the interests and experiences of specific sex affiliation, but less significantly associated with intelligence.

In comparing the difference of play activities most preferred by the lowest and highest levels, it may be well to list first the play activities which appeared in outstanding choices of the lowest mental level but did not appear in these choices of the highest mental level, and then list the play activities which appeared in the highest mental level but did not appear in the lowest mental level (see Tables 2 and 3).

It appears evident that both girls and boys in the highest mental level were more likely to favor play activities which demand more intelligence or some basic skills, acquired in the school or in the cultural society of which they are members. This certainly is apparent when we consider certain activities such as play tag, hide and seek, drop the handkerchief, playing jacks, and

TABLE 2

PLAY ACTIVITIES WHICH APPEARED MOST FREQUENTLY IN THE LOWEST BUT NOT THE HIGHEST MENTAL LEVELS

(Girls)	(Boys)
1. Play tag	1. Play tag
2. London Bridge	6. Hide and seek
4. Farmer in the dell	
5. In and out the windows	
6. Hide and seek	
8. Drop the handkerchief	
26. Play jacks	

TABLE 3

PLAY ACTIVITIES WHICH APPEARED MOST FREQUENTLY IN THE HIGHEST BUT NOT THE LOWEST MENTAL LEVELS

(Girls)	(Boys)
12. Dodge ball	10. Postoffice
23. Play with dolls	24. Play with top airplanes
35. Hike, or take long walks	93. Go to the movies
37. Dance	
42. Swim	
65. Play piano, violin, etc.	
67. Read books and stories	
93. Go to the movies	

London Bridge, all preferred with greater frequency among the low intelligence girls, and contrast them with activities such as playing with toy airplanes, playing musical instruments, and reading. Withal, though, one of the most evident features of the plays and games preferred with greater frequency by superior than by dull children is the greater breadth of types of play enjoyed by the superior group. They have a rather fine balance of interests, including both distinctly physical and intellectual pursuits in their list of greater preferred games.

The play activities that are shown in Tables 4 and 5 were selected on the basis of having a critical ratio of as much as 3.00 between the percentages of preference of the lowest and highest intelligence groups. In the main, we see the same, or at least the same types of activities which were noted in the previous analysis of differences in play preferences of bright and dull children. Although only about 17 or 18 statistically significant differences out of a possible 216 have been discovered, it is worthy of note that this is a more significant relationship than was found between play interests and any of the other variables studied (2, 3).

One thing to which the writer should like to call attention is what seems to be something of an apparent trend for girls of higher mental level to like

TABLE 4
PLAY ACTIVITIES WHICH WERE FAVORED MOST FREQUENTLY BY CHILDREN IN THE LOWEST GROUP

Play activities (Girls)	t	Play activities (Boys)	t
3. London Bridge	7.37	8. Drop the handkerchief	5.44
4. Farmer in the dell	7.00	1. Play tag	4.65
8. Drop the handkerchief	5.91	6. Hide and seek	3.34
5. In and out the windows	5.89		
6. Hide and seek	5.39		
2. Ring around the rosy	4.73		
9. Blindfold	4.16		
1. Play tag	3.93		
80. Play "Old Maid"	3.65		
11. Cat and mouse	3.05		

TABLE 5
PLAY ACTIVITIES WHICH WERE FAVORED MOST FREQUENTLY BY CHILDREN IN THE HIGHEST GROUP

Play activities (Girls)	t	Play activities (Boys)	t
42. Swim	6.06	10. Postoffice	6.11
93. Go to the movies	5.98	63. Football	5.11
67. Read books and stories	4.85	60. Tennis	4.83
97. Take Kodak pictures	4.55	93. Go to the movies	4.81
28. Walk on stilts	3.52	24. Play with toy airplanes	3.30
65. Play piano, violin, etc.	3.38		
32. Ride bicycle	3.24		
12. Dodge ball	3.18		
37. Dance	3.09		
43. Ride horseback	3.02		

to engage in play activities which are somewhat quiet and esthetic in nature, and for boys of this type to like to engage in play activities that call for motor skill, foresight, and romantic appreciation.

2. *An Analysis of Play Activities Disliked Most Frequently by Children at Different Mental Ability Level*

Table 6 shows that the percentage values of play activities 2, 3, 13, 23, and 49 for girls, and 2, 3, 4, 7, 23, 36, and 37, for boys, increase consistently or progressively with increasing mental ability scores. Ring around the rosy, London Bridge, and playing with dolls were disliked more by both girls and boys who ranked highest in mental ability than those who ranked lowest in mental ability. The results with respect to activities 2, 3, are in perfect accord with the results found in the last section.

Girls of highest mental ability disliked puss in the corner and leap-frog. Boys of highest mental ability appear to dislike play activities which may be

TABLE 6

A Comparison of Certain Play Activities Disliked Most by at Least 10 Per Cent of the Children in Any One of the Three Mental Ability Groups

Play activities	Percentages of participation Low.	Mid.	High	*PE* of % Low.	Mid.	High
		(*Girls*)				
1. Play tag	15.5	15.9	15.2	2.5	1.8	2.2
2. Ring around the rosy	16.5	32.2	35.6	2.6	2.4	3.0
3. London Bridge	11.4	16.4	17.0	2.2	1.8	2.4
4. Farmer in the dell	6.2	12.6	9.3	1.7	1.6	1.8
5. In and out the windows	7.2	11.0	9.3	1.8	1.6	1.8
7. Hopscotch	9.3	11.0	5.9	2.0	1.6	1.4
8. Drop the handkerchief	6.2	12.6	4.2	1.7	1.6	1.2
10. Postoffice	14.4	13.1	12.7	2.4	1.7	2.0
11. Cat and mouse	13.2	9.3	5.9	2.3	1.5	1.4
13. Puss in the corner	5.2	10.2	11.8	1.6	1.6	2.0
15. Marbles	20.6	19.7	11.0	2.8	2.0	2.0
20. Play with fire	33.0	23.0	30.5	3.2	2.2	2.9
23. Play with dolls	6.2	7.6	13.6	1.7	1.3	2.1
24. Play with toy airplanes	10.3	7.6	8.5	2.1	1.3	1.7
26. Play jacks	10.3	6.0	6.8	2.1	1.2	1.5
27. Spin tops	6.2	6.0	11.8	1.7	1.2	2.0
49. Leap-frog	8.2	11.5	15.2	1.9	1.6	2.2
		(*Boys*)				
1. Play tag	17.5	7.7	9.7	2.5	1.3	2.0
2. Ring around the rosy	21.4	33.1	39.7	2.7	2.5	3.2
3. London Bridge	13.6	21.9	24.2	2.3	2.2	2.8
4. Farmer in the dell	16.5	16.6	22.3	2.4	2.0	2.8
5. In and out the windows	12.6	18.9	12.6	2.2	2.0	2.2
7. Hopscotch	9.7	11.8	20.4	2.0	1.7	2.7
10. Postoffice	14.6	9.5	10.7	2.4	1.5	2.0
11. Cat and mouse	3.9	10.1	5.8	1.1	1.6	1.5
13. Puss in the corner	12.6	7.7	5.8	2.2	1.3	1.5
20. Play with fire	18.4	17.8	11.6	2.6	2.0	2.1
23. Play with dolls	13.6	24.2	26.2	2.3	2.2	3.0
25. Roll hoops	6.8	10.1	5.8	1.6	1.6	1.5
26. Play jacks	10.7	9.5	19.4	2.0	1.5	2.6
36. Do garden work	2.9	5.9	15.5	0.9	1.1	2.4
37. Dance	7.8	8.9	14.6	1.8	1.4	2.4
48. Jump the rope	9.7	9.5	12.6	2.0	1.5	2.2
49. Leap-frog	9.7	8.3	13.6	2.0	1.4	2.3

classified into the following categories: (*a*) physical labor, such as doing garden work; (*b*) conventionally considered feminine or "sissy" types of play activities, such as playing dolls or dancing; (*c*) immature types of play, such as ring around the rosy, London Bridge, farmer in the dell, and hopscotch.

It may be seen from Table 6 that the percentage values of play activities 10, 11, and 15, for girls, and 13 and 20, for boys, have a consistent or progressive tendency to decrease as mental ability scores increase, or conversely, the amount of dislike for these activities increases as mental ability scores decrease.

To be sure, some of the differences are very small, but there is some general agreement with the finding in the last section. On the other hand, activities such as 11 for girls, and 13 for boys may seem to be associated inconsistently with intelligence. The apparent discrepancy probably was due to the somewhat evident tendency for superior children to ignore these activities when mentioning their dislikes.

Play activities 2, ring around the rosy, 3, London Bridge, and 20, playing with fire, were among the most disliked by both girls and boys in every mental level. Activity 15, marbles, was disliked by girls of every mental level, and it seems obvious that playing of marbles reveals a definite influence of sex affiliation, or definite evidence of how experiences, mores, and social opportunities affect play interests. Play activities 4, 5, and 23 were among the most disliked by boys in every mental level, and the same interpretation given to the dislike of marbles by girls also may be given to this instance for the boys. It may be of some interest to note that a general trend for both girls and boys to decrease the degree of their disliking attitude toward *activity 10, postoffice, as their mental ability scores increase, seem to have existed, as indicated in Table 6. However, this tendency seems a little obscure, and no one can say with assurance that this tendency is truly affected* by intelligence as a major factor.

Activity 49, leap-frog, was disliked more by both girls and boys of the highest mental group than by children of either the lowest or middle mental group. The lowest mental ability girls disliked to play jacks almost twice as frequently as the highest mental girls, and on the other hand, the highest mental ability boys disliked to play the same game almost twice as frequently as did the lowest mental boys. The reason why they have these differences is not apparent.

Among these 14 play activities, as shown in Table 7 and Table 8, there are only five whose differences are statistically significant, and four of these five show a greater incidence of dislike among the highest mental group. Activity 2, ring around the rosy, was much more disliked by both girls and boys of the superior mentality than by children of the inferior mentality.

TABLE 7
PLAY ACTIVITIES DISLIKED MORE BY CHILDREN IN THE LOWEST GROUP

Play activities (Girls)	*t*	Play activities (Boys)	*t*
35. Hike, or take long walks	3.10	16. Duck on rock	4.09
29. Row a boat	3.10	27. Spin tops	3.67
		24. Play with toy airplanes	3.24
		12. Dodge ball	3.01

TABLE 8

Play Activities Disliked More by Children in the Highest Group

Play activities (Girls)	*t*	Play activities (Boys)	*t*
2. Ring around the rosy	4.91	36. Do garden work	4.91
70. Checkers	4.54	2. Ring around the rosy	4.37
57. Running races	3.14	23. Play with dolls	3.33
		65. Play piano, violin, etc.	3.24
		7. Hopscotch	3.16

By examining the list as a whole, it seems quite obvious that, to say the least, mentality is not the major determinant affecting children's dislikes of these activities. As has been suggested, individual experiences, mores, and social opportunities probably have combined influences upon the child in controlling his expression of his dislikes.

In Table 9 one will notice a summary of the number of play activities enjoyed by children of varying degrees of tested mental ability. A careful analysis of this table will reveal the lack of consistency of any trends or apparent trends which may be seen. Thus, though middle boys appear to have a slightly greater breadth of play interests than boys of either intelligence extreme, we find that among girls the children of middle or average tested intelligence have a slightly lower number of play interests, on the average, than is found for either of the extreme groups.

TABLE 9

Analysis of Number of Play Activities Enjoyed by Children of Designated Mental Abilities

Sections on basis of mental ability test scores	*Mean number of play activities enjoyed*		*Sigma*		*PE_m*	
	Girls	Boys	Girls	Boys	Girls	Boys
All children	52.6	51.2	19.9	21.2	0.44	0.48
Lowest 10%	56.0	50.4	17.9	22.2	1.22	1.48
Middle 20%	52.3	53.6	20.0	22.0	1.00	1.15
Highest 10%	53.2	50.5	17.5	20.4	1.08	1.36

In furtherance of the idea that mental ability is not a factor of major significance in the determination of breadth of play interests, it should be noted that, within each sex group, the difference between the mean number of play activities of the total group and the mean number of any designated intelligence group is less than three times its probable error, and hence at best of no more than questionable reliability. Not only are there no significant differences between specified group means, but also there is no significant difference in homogeneity of play interests as reflected in the standard deviations of the groups.

D. SUMMARY

1. Swimming and going to movies are the only ones of the 108 activities about which there is a progressively increasing expression of interest for both sexes along with increasing intelligence, as this study revealed.

2. In general, with increases in brightness there is something of an increasing interest in play of a rather active nature such as hiking, baseball, and bicycle riding, but a decreasing interest in immature forms of recreation such as playing tag, hide and seek, and drop the handkerchief.

3. It appears rather evident that both girls and boys at the highest mental level are somewhat more likely to favor play activities which demand more intelligence or some basic skills, acquired in the school or in the cultural society of which they are members. Girls of higher mental level seem to like more the activities which are somewhat sedentary, solitary, and esthetic in nature than do the dull girls, and boys of higher mental level tend to enjoy more the activities that call for motor skill, initiative, foresight, and romantic appreciation.

4. Most of the children at the lowest mental level are likely to enjoy play activities which have no complicated rules, or which require just the amount of intelligence that a child may have to understand the ways of playing.

5. Ring around the rosy is much more disliked by both girls and boys of the highest mentality than by those of the lowest mentality. Girls of the highest mentality dislike much more to play checkers than do girls of the lowest mentality, and boys of the highest mentality dislike much more to "do garden work" than do boys of the lowest mentality.

6. Mental ability may be a factor but it certainly is not the factor of major significance in the determination of breadth of play interests.

REFERENCES

1. BOYNTON, P. L., & FORD, F. A. The relationship between play and intelligence. *J. Appl. Psychol.*, 1933, **17**, 294-301.
2. BOYNTON, P. L., & WANG, J. D. Relation of the play interests of children to their economic status. *J. Genet. Psychol.*, 1944, **64**, 129-138.
3. ———. Relationship between children's play interests and their emotional stability. *J. Genet. Psychol.*, 1944, **64**, 119-127.
4. DIMOCK, H. S. Rediscovering the Adolescent. New York: Associated Press, 1937. Pp. 287.
5. HAGGERTY, M. E., & NASH, H. B. Mental capacity of children and parental occupation. *J. Educ. Psychol.*, 1924, **15**, 559-572.

6. Hollingworth, L. S. Gifted Children. New York: Macmillan, 1927. Pp. 374.
7. Lehman, H. C., & Witty, P. A. A study of play in relation to intelligence. *J. Appl. Psychol.*, 1928, **12**, 369-397.
8. Lewis, W. D. A study of superior children in the elementary school. *Geo. Peabody Coll. Contrib. Educ.*, 1939, **246**. Pp. 104.
9. McGehee, W. A study of retarded children in the elementary school. *Geo. Peabody Coll. Contrib. Educ.*, 1939, **246**. Pp. 128.
10. Terman, L. M. Genetic Studies of Genius, Vol. 1. Stanford University: Stanford Univ. Press, 1925. Pp. 648.
11. Wang, J. D. A study of certain factors associated with children's play interests. *Geo. Peabody Coll. Contrib. Educ.*, 1941, **248**. Pp. 123.
12. Witty, P. A., & Beaman, F. N. The play of mental deviates. *Ment. Hyg.*, 1933, **17**, 618-634.
13. Witty, P. A., & Lehman, H. C. The play behavior of fifty gifted children. *J. Educ. Psychol.*, 1927, **18**, 259-265.

Galesburg State Research Hospital
Galesburg, Illinois

The Journal of Genetic Psychology, 1958, **93**, 133-148

AN ANALYSIS OF EXPRESSIONS USED BY ADOLESCENTS AT VARYING ECONOMIC LEVELS TO DESCRIBE ACCEPTED AND REJECTED PEERS*

Baruch School of Business and Public Administration, City College of New York
Mergenthaler Linotype Company

MORTIMER R. FEINBERG, MAX SMITH, AND ROBERT SCHMIDT

A. PROBLEM

A number of studies have been concerned with the dynamics of sociometric choice at the adolescent level. The problems undertaken have ranged from identification of the reasons for the selection pattern (1, 3, 4, 5) to the quantification of background experience associated with acceptance or rejection (2).

Some studies have attempted to analyze the reasons given by adolescents themselves for their choices. Other more ambitious projects have tried to determine whether these descriptive statements differed at varying age levels (6). However, we know of no previous research which has attempted to compare these statements at contrasting economic levels.

The problem in this investigation was to determine whether adolescent boys at three different economic levels would use similar descriptive terms to characterize the peers they accepted or rejected. Which, if any, characteristics would be common to all adolescents and which would be specific to a particular economic level? This study attempted to find the answers to these questions. The results should be of interest to social class researchers, to analysts of adolescent behavior, and to educators.

B. SUBJECTS AND PROCEDURES

The sample consisted of over 2,400 adolescent boys between the ages of 13 and 15, drawn from schools representing three different family income levels in New York City and its suburbs. The same sample had been used in an earlier study concerned with the relationship between background experience and social acceptance (2). Table 1 describes the levels and numbers of boys included.

A large *N* was required to insure sufficient frequencies for the comparison in the use of descriptive terms among the different economic levels. This

---

*Received in the Editorial Office on June 17, 1957.

TABLE 1
DESCRIPTION OF SAMPLE OF ADOLESCENT BOYS, AGES 13 THROUGH 15 (IN 1948)

Level	Number of boys	Schools
High—family incomes at least $10,000	1,050	31 classrooms in 8 private schools, Westchester County, N. Y.
Middle—family incomes between 5,000-$7,500 per year	775	19 classrooms in 3 public schools, Bronx, N. Y.
Low—family incomes under $3,000 per year	650	21 classrooms in 5 public schools, Manhattan, N. Y.

(Total $N = 2{,}475$.)

prohibited the collection of intricate data on each individual's socioeconomic status.

There were, moreover, prominent ethnic differences associated with the income areas chosen, so that the results could probably be reported in these terms. The high income level boys were predominantly Protestant; the middle income boys, largely Jewish; and the low income groups, Catholic—mainly Ialian and Irish. However, the primary discriminator was income level, and results will be reported in these terms.

The high economic-level sample was drawn from private schools charging high tuition rates. The students came from the "Gold Coast" or wealthy suburban areas where the family incomes are estimated (by the headmasters of these schools) to be at least $10,000 and, in most cases, much greater.

The middle and low economic levels were taken from the public schools of New York City by first consulting an official of the Board of Education who assigned an economic level to the different schools on the basis of neighborhood, reports from census figures, etc. Then the principals of the pacticular schools classified their own students as belonging in the main to a particular economic level. There was complete agreement between the officials' and the principals' estimates. While there may be a few students in each of the selected schools who do not belong to the economic level of the majority of their classmates, their number is probably insignificant.

The authors make no claim that the levels represented here have equivalence to any of the usual social class levels referred to so frequently these days; it would be presumptive to claim such equivalence, especially for a sample drawn from the complex socioeconomic structure of Metropolitan New York. Nevertheless, these groups are differentiated on the basis of economic level and, for simplicity, will be referred to in this paper as high, middle, and low income groups.

The pupils at each level were asked to list adjectives which in their opinion were descriptive of the four boys in their class whom they would like to have sitting next to them the following term—who in their opinion would make the most desirable friends; and also the adjectives descriptive of the four boys with whom they would feel uncomfortable, who gave them a feeling of uneasiness or annoyance in the classroom situation. This part of the data collection was accomplished during the earlier study (2).

For the present study, the first tabulation included the number of times each term was used. Then we grouped into acceptant-term clusters and rejective-term clusters the expressions that seemed to belong together.[1] Finally we combined into "cluster dichotomies" term clusters of opposite polarity—that is, those which in our opinion really represented opposite poles of the same general personality characteristics. The dichotomizing, though at times rather arbitrary, seems to make the data more intelligible.

C. Data Treatment

We were interested in differences in the values held by the three economic groups. Raw frequency differences among these groups are due also, however, to differences in (*a*) the average number of peers, (*b*) the average number of terms used to characterize accepted or rejected peers in each group ("verbosity"), and (*c*) the size of the groups.

Chi-square was used to determine which clusters differentiate significantly among the three socio-economic classes.

First, while keeping constant the *grand total* of acceptant-term frequencies, the raw frequency of each acceptant-term cluster was adjusted so as to equate for the effects of the following two factors among the economic groups: (*a*) the average number of accepted peers named by each subject in the group, (*b*) the average number of terms used in the group for each accepted peer (acceptant verbosity). Likewise keeping constant the grand total of rejective-term frequencies, each raw rejective frequency was adjusted to equate for the average number of rejected peers named by each subject in the group and for the average rejective verbosity of the group. For each acceptant-term cluster and for each rejective-term cluster Chi-square, based on these adjusted frequencies, was used to decide whether the differences found among the groups were significant at the one per cent level.

The frequencies were then further adjusted for differences in size of group. The numbers in Columns 6, 7, and 8 of Table 3 are these adjusted

---

[1] The instances in which favorable terms were applied to rejected peers, or unfavorable terms applied to accepted peers, were negligible (fewer than one instance per 200 peers described). These instances were disregarded in the analysis.

TABLE 2
CLUSTER DICHOTOMY

Identifying Number	Name	Illustrative Terms
1	Intelligent	Intelligent, keen, bright, etc. vs. Dumb, stupid, etc.
2	Sociable	Friendly, sociable, etc. vs. Unfriendly, too quiet, stiff, etc.
3	Minds own business	Minds own business, etc. vs. Annoying, pest, etc.
4	Plays fair	Good sport, plays fair, etc. vs. Poor sport, etc., plays unfair
5	Quiet	Quiet vs. Loud, noisy, overtalkative, etc.
6	Witty	Humorous, witty, good joker, etc. vs. Not humorous, etc., no sense of humor
7	Athletic	Athletic, ball player, etc. vs. Not athletic, etc.
8	Helpful	Helpful vs. Not helpful
9	Unconceited	Humble, doesn't show off, etc. vs. Conceited, stuck up, etc.
10	Good company	Good company, fun to be with, etc. vs. No fun, poor company, etc.
11	Serious	Serious, not silly, etc. vs. Silly, foolish, etc.
12	Conscientious	Conscientious, good worker, etc. vs. Lazy, listless, etc.
13	Masculine	Real man, has guts, etc. vs. Sissy, helpless, girlish, fairy
14	Stays out of trouble	Stays out of trouble vs. Always in trouble, etc.
15	Talks well	Can talk, talks well, vs. Can't talk
16	Honest	Honest, doesn't cheat, vs. Dishonest, cheats, lies, etc.
17	Clean	Clean, neat, vs. Sloppy, dirty, etc.
18	Doesn't fight	Doesn't fight vs. Always fighting, etc.
19	Kind	Kind, considerate, vs. Unkind, not considerate, mean
20	Trustworthy	Trustworthy, keeps his word, etc. vs. Unreliable, etc.
21	Gets along well with others	Gets along well wih others, etc. vs. Can't get along
22	Leader	Leader vs. Not a leader
23	Cheerful	Cheerful, etc. vs. Grumpy, complains, etc.

TABLE 2 (*continued*)

Identifying Number	Name	Illustrative Terms
24	Coöperative	Coöperative vs. Not coöperative
25	Good scholar	Good scholar, student, etc. vs. Poor scholar, student, etc.
26	Common interests	Same interests vs. Not same interests
27	Interesting	Interesting vs. Not interesting, dull, etc.
28	Good manners	Good manners vs. Poor manners
29	Pleasant, agreeable	Pleasant, agreeable vs. Argues, insults, etc.
30	Can take a joke	Can take a joke vs. Can't take a joke
31	Mature	Mature, grownup, etc. vs. Immature, babyish
32	Generous	Generous, unselfish vs. Tight, selfish
33	Good-looking	Good-looking, clean cut, etc. vs. Ugly, etc.
34	Good Character	Good character vs. Poor character
35	Understanding	Understanding vs. Not understanding
36	Participates in activities	All around vs. No activities, doesn't take part, etc.
37	Calm	Calm, doesn't get excited, easy-going vs. Bad tempered, gets excited, etc.
38	Sincere	Sincere, means what he says, etc. vs. Insincere
39	Well dressed	Good dresser, sharp clothes, etc. vs. Poor dresser
40	Other specific terms used infrequently	
41	General non-specific terms	Swell, good friend vs. No good, real drip

frequencies.[2] They permit direct comparisons among the groups and form the basis for the conclusions drawn. The data were also analyzed without any adjustment for relative verbosity; but this analysis will not be presented, since it did not materially change any of the conclusions.

---

[2] *Method used to adjust the raw frequencies.* Altogether the 2,475 subjects used 5,869 acceptant terms. Had every subject accepted the same number of peers and used the same number of terms for each, each subject would have used 1/2,475 of the 5,869 acceptant terms, or 2.371 acceptant terms. If we assume (*a*) that each group had 1,000 individuals, (*b*) that all groups were equal in "acceptant verbosity,"

Some term clusters were used too infrequently to warrant the application of the Chi-square test, or, in general, intergroup comparisons. A good deal of subjective judgment was involved in categorizing the 8,751 terms. We ended up with 39 dichotomized categories, but might easily have come out with more or perhaps fewer. We decided arbitrarily that unless the adjusted average number of mentions was more than 20 per thousand cases for any acceptant-term cluster or rejective-term cluster in these 39 categories, statistical intergroup comparisons would not be warranted for that cluster.

D. FINDINGS

1. *Specific vs. General Terms*

A number of the terms used could not be placed into any of the specific term categories; they were merely general expressions of approval such as "swell" or "wonderful guy," or of disapproval such as "no good" or "real drip." Among the acceptant terms, 27 per cent were of this general nature, whereas among the rejective terms only 3 per cent were. Apparently acceptance is its own excuse for being, whereas rejection would seem to demand a more detailed self-justification.

2. *Within-Group Values*

For convenience, each cluster dichotomy will be referred to as a trait. However, this is a very loose use of the term and it refers only to the adjective descriptions, not to the strict definition of the term as used in personality theory.

In the high group, which traits lead to acceptance, which to rejection? Which traits lead to acceptance or rejection in the middle group? Which in the low group? To help clarify the picture of the values at each income level, the descriptive terms considered important by each group were classified into the following three categories:

1. *Two-way traits.* Boys tend to focus their attention upon both the positive and the negative aspects of these characteristics when discriminating

---

and (*c*) that the average acceptant verbosity per subject remained 2.371 acceptant terms, then each group would have used a total of 2,371 acceptant terms.

In the high group, however, the number of acceptant terms actually used was 2,685. If each term frequency is multiplied by 2,371/2,685, or 0.883, we shall come out with the desired total of 2,371. So 0.883 is our adjustment factor for acceptant terms in the high group. Similarly, we obtain for the mid group an acceptant adjustment factor of 2,371/1,462, or 1.662, and for the low group, 2,371/1,722, or 1.377.

Average rejective verbosity was 2,882/2,475, or 1.164 rejective terms per subject. For the high group, then, the rejective adjustment factor is 1,164/1,417, or 0.821; for the mid group, 1,164/805, or 1.446; and for the low group, 1,164/ or 1.764.

TABLE 3

Frequency of Acceptant and Rejective Term Clusters in the Three Socio-Economic Groups

Identifying number	Acceptance or rejection	Raw frequency: High (1,050)	Raw frequency: Middle (775)	Raw frequency: Low (650)	Frequency per 1,000 after adjustment for number of peers named and for verbosity: High (1,000)	Middle (1,000)	Low (1,000)	Significance[a]
1	A	263	75	77	232	122	106	S[b]
	R	201	14	11	165	20	19	S
2	A	307	79	69	271	128	95	S
	R	161	24	7	132	35	12	S
3	A	1	35	46	1	57	63	S
	R	72	162	112	59	234	198	S
4	A	76	87	199	67	141	274	S
	R	22	11	30	18	16	53	S
5	A	42	57	76	37	92	105	S
	R	57	85	84	47	123	148	S
6	A	184	85	92	163	138	127	N[c]
	R	14	16	0	12	23	0	—[d]
7	A	227	37	85	201	60	117	S
	R	48	0	0	39	0	0	—
8	A	63	78	128	56	126	176	S
	R	2	3	5	2	4	9	—
9	A	6	6	0	6	10	0	—
	R	117	92	69	96	133	122	N
10	A	56	85	92	49	138	127	S
	R	5	0	5	4	0	9	—
11	A	18	0	0	16	0	0	—
	R	61	83	66	50	120	117	S
12	A	113	23	22	100	39	30	S
	R	20	28	6	16	41	11	—
13	A	15	3	0	13	5	0	—
	R	127	36	20	104	52	35	S

between accepted and rejected peers. Thus, for example, accepted peers are very often characterized as friendly, and rejected peers, also frequently, as unfriendly.

2. *"Bonanza" traits.* Apparently attention is focused almost entirely on the positive aspect of these characteristics. Their negative aspect tends to be disregarded. Consequently, the positive aspect is frequently associated with acceptance, whereas the negative aspect is seldom mentioned at all. Thus, accepted peers are often designated as humorous, witty, a good joker, etc., while very seldom is anyone characterized as not being witty. We are probably justified in assuming that the presence of a "bonanza" trait fosters acceptance but that its absence is not a cause for rejection.

3. *"Blackball" traits.* This category, the opposite of the bonanza group, includes traits that seem important mainly for rejection. Here attention seems to be centered primarily on the negative aspect of the trait. Thus, rejected peers are frequently seen as conceited, stuck-up, whereas individuals are seldom characterized as "not conceited." Traits of this nature we shall term "blackball" traits.

Table 4 lists the two-way traits, the bonanza traits, and the blackball traits for each of the three economic groups.

3. *Intergroup Comparisons*

In the previous section our interest was focused upon the bases for acceptance or rejection *within* each of the three economic groups. Here we are concerned with a comparison among the three groups.

a. Universals. Certain term clusters are valued in a more or less similar fashion by all three groups. These, within the context of this study, may be labeled *universals*. The terms themselves are listed in the upper part of Table 5.

b. Differentiating features. With respect to the remaining term clusters, though, the peer-values of the high economic group are in many respects quite different from those of the two lower levels. On the other hand, the peer-values of the middle and low income groups resemble each other's closely. Table 5 shows in detail how the bases for acceptance or rejection compare in the three economic groups.

E. DISCUSSION

Tapping adolescent opinion through descriptions of accepted and rejected peers has revealed some interesting findings. It is, of course, undesirable to generalize too far from any one study.

TABLE 3 (*continued*)

Identifying number	Acceptance or rejection	Raw frequency High (1,050)	Raw frequency Middle (775)	Raw frequency Low (650)	Frequency per 1,000 after adjustment for number of peers named and for verbosity High (1,000)	Frequency per 1,000 after adjustment for number of peers named and for verbosity Middle (1,000)	Frequency per 1,000 after adjustment for number of peers named and for verbosity Low (1,000)	Significance[a]
14	A	0	7	10	0	11	14	—
	R	13	48	58	11	69	102	S
15	A	22	50	44	19	81	61	S
	R	0	0	0	0	0	0	—
16	A	52	17	29	46	28	40	N
	R	26	5	10	21	7	18	—
17	A	20	14	27	18	23	37	N
	R	35	11	18	29	16	32	N
18	A	0	7	12	0	11	17	—
	R	26	61	74	21	88	131	S
19	A	58	7	43	51	11	59	S
	R	42	0	13	35	0	23	—
20	A	35	6	25	31	10	35	S
	R	22	13	18	18	19	32	—
21	A	31	31	18	27	50	25	S
	R	0	12	6	0	17	11	—
22	A	108	0	0	95	0	0	S
	R	28	0	0	23	0	0	—
23	A	85	9	10	75	15	14	S
	R	7	4	1	6	6	2	—
24	A	48	13	11	42	21	15	S
	R	27	0	0	22	0	0	—
25	A	57	0	8	50	0	11	—
	R	12	12	0	10	17	0	—
26	A	3	36	24	3	59	33	S
	R	0	0	0	0	0	0	—

TABLE 3 (*continued*)

Identifying number	Acceptance or rejection	Raw frequency High (1,050)	Raw frequency Middle (775)	Raw frequency Low (650)	Frequency per 1,000 after adjustment for number of peers named and for verbosity High (1,000)	Frequency per 1,000 after adjustment for number of peers named and for verbosity Middle (1,000)	Frequency per 1,000 after adjustment for number of peers named and for verbosity Low (1,000)	Significance[a]
27	A	24	15	0	21	24	0	—
	R	37	6	0	30	9	0	—
28	A	22	26	11	19	42	15	S
	R	6	0	2	5	0	4	—
29	A	26	25	14	23	41	19	S
	R	27	2	2	22	3	4	—
30	A	2	10	22	2	16	30	—
	R	1	4	6	1	6	11	—
31	A	17	0	0	15	0	0	—
	R	59	0	0	49	0	0	—
32	A	18	9	15	16	15	21	—
	R	0	7	0	0	10	0	—
33	A	26	0	12	23	0	17	—
	R	15	0	2	12	0	4	—
34	A	10	3	8	9	5	11	—
	R	0	18	2	0	26	4	—
35	A	11	5	11	10	8	15	—
	R	0	0	0	0	0	0	—
36	A	32	0	0	28	0	0	—
	R	30	0	0	25	0	0	—
37	A	10	0	0	9	0	0	—
	R	10	5	0	8	7	0	—
38	A	22	0	0	19	0	0	—
	R	4	0	0	3	0	0	—
39	A	6	3	5	5	5	7	—
	R	4	0	1	3	0	2	—

TABLE 3 (*continued*)

Identifying number	Acceptance or rejection	Raw frequency			Frequency per 1,000 after adjustment for number of peers named and for verbosity			
		High (1,050)	Middle (775)	Low (650)	High (1,000)	Middle (1,000)	Low (1,000)	Significance[a]
40[e]	A	42	0	32	37	0	44	—
	R	59	11	8	49	16	14	—
41	A	527	519	445	465	842	613	—
	R	20	32	24	16	46	42	—
Total	A	2,685	1,462	1,722	2,370	2,374	2,373	—
	R	1,417	805	660	1,163	1,163	1,169	—

[a] After adjustment for number of peers named and for verbosity, but using the actual number of subjects in each group.
[b] Differences are significant at the 1 per cent level.
[c] Differences are not significant at the 1 per cent level.
[d] Total frequency is not considered high enough to warrant testing for significance of the differences.
[e] Significance figures for Categories 40 and 41 would be pointless in view of the broadness of these categories.

TABLE 4

TRAITS CONSIDERED IMPORTANT AT VARIOUS ECONOMIC LEVELS

	High	Middle	Low
Two-way[a]	Intelligent *vs.* stupid (1) Sociable *vs.* unfriendly (2) Quiet *vs.* Noisy (5) Athletic *vs.* not athletic (7) Honest *vs.* dishonest (16) Kind *vs.* mean (19) Coöperative *vs.* not coöperative (24)	Mind own business *vs.* pest (12) Quiet *vs.* noisy (5) Conscientious *vs.* lazy (12)	Sociable *vs.* unfriendly (2) Minds own business *vs.* pest (3) Plays fair *vs.* plays unfair (4) Quiet *vs.* noisy (5) Kind *vs.* mean (19) Trustworthy *vs.* unreliable (20)
Bonanza[b]	Plays fair (4) Witty (6) Helpful (8) Conscientious (12) Good company (10) Gets along well with others (22) Leader (22) Cheerful (23) Good Scholar (25)	Intelligent (1) Sociable (2) Plays fair (4) Witty (6) Athletic (7) Helpful (8) Good company (10) Talks well (15) Gets along well with others (21) Common interests (26) Good manners (28) Pleasant, agreeable (29)	Intelligent (1) Witty (6) Athletic (7) Helpful (8) Good company (10) Conscientious (12) Talks well (1) Common interests (26)
Blackball[c]	Pest (3) Conceited (9) Silly (11) Sissy (13) Immature (31) No activities (36)	Conceited (9) Silly (11) Sissy (13) Always in trouble (14) Always fighting (18)	Conceited (9) Silly (11) Sissy (13) Always in trouble (14) Always fighting (18)

[a] *Two-way.* Attention is paid to both the positive and negative aspects.

[b] *Bonanza.* Concentration is just on the positive aspects. The negative aspects are disregarded.

[c] *Blackball.* Concentration is just on the negative aspects of the trait.

TABLE 5
DIFFERENTIAL SIGNIFICANCE OF TRAITS IN VARIOUS ECONOMIC GROUPS

	Positive traits	Negative traits
A. Universals* Important at all levels—high, middle, low economic levels.	Intelligent (1) Plays Fair (4) Quiet (5) Athletic (7) Good company (10) Conscientious (12) Honest (16) Can take a joke (30)	Pest (3) Loud, noisy (5) Conceited (9) Silly (11) Sissy, girlish (13)
B. Universals Relatively more important to the high group.	Intelligent (1) Athletic (7) Conscientious (12)	
C. Traits important to the high group only.	Leader (22) Cheerful (23) Coöperative (24) Good scholar (25) Participates in activities (36)	Stupid (1) Not athletic (7) Not a leader (22) Immature (31) No activities (36)
D. Universals Relatively more important in the middle and low groups.	Good sport (4) Quiet (5) Good company (10)	Pest (3) Noisy (5) Silly (11)
E. Traits important to the middle and low groups, but not to the high group.	Minds own business (3) Talks well (15) Common interests (26)	Always fighting (18)
F. One trait that seems important to the high and low group but not to the middle group.	Kind (19)	Mean (19)

*Universals—Traits valued similarly by all economic groups.

Despite the large sample there are some obvious limitations to the conclusions. The concentration of this sample was in the greater New York area. Moreover, even within this restricted population, certain environmental variables were uncontrolled. The high socio-economic group attended private schools. Would peer acceptance values of a control group consisting of equally wealthy boys attending public schools be the same? The high group was in the main, Protestant; the middle group, Jewish; the low group, Catholic. These religious differences may account for some of the variance which we have attributed to the income disparity. The results are also confined to boys within the 13-15 age bracket. Other age groups may have different values. However, in spite of these obvious limitations, there are still many important things to be learned from the use of descriptive phrases supplied by adolescents.

1. Boys of 13 to 15 pick their accepted peers for the "right" reasons—because of certain desirable personality and background traits. The present study shows that they reach out and desire close contact with peers who are described as "intelligent," "fair," "good company," "athletic," "conscientious," and "honest."

This finding corroborates an eariler study with the same subjects, which analyzed their background experiences and concluded that accepted boys, at all economic levels, appear to have made an adequate adjustment at school and at home. They were proficient in both scholastic and athletic activities, and their parents were pleased with their grades. Also, in the opinion of the adolescents, the parents had no favorites in the family group (2). A more recent study found that similar background factors were correlated with election to offices in a junior high school (5).

Guidance counselors working with typical adolescents, even at low economic levels, might be reassured by these findings.

2. Rejected adolescents are by-passed for reasons that are likewise good and substantial. They are seen by their classmates as pests, loud, noisy, silly, and conceited.

This finding, too, agrees with those of the background study, which showed that rejected adolescents at all economic levels are less proficient in athletic and scholastic activities, have poor relations with teachers, are more withdrawn, and have fewer close friends.

Mental hygiene researchers and therapists might well direct more of their efforts to studying the rejected youngster. Attention to the fact of rejection at this age level, when personality may be a little more pliable than later, should pay rich dividends in subsequent individual adjustment. The present research suggests that unpopularity cannot be lightly dismissed as a "passing phase."

3. The middle income group and the low income group display a basic similarity in values, at least in choice patterns, which sets them apart from their high economic counterparts. In only one respect—considering kindness as important for acceptance and meanness as a basis for rejection—did the low group resemble the high-income but not the middle-income group.

Of special interest is the middle and low income level adolescents' stress on the verbalization ability. The popular adolescent at these two levels seems to possess verbal fluency. It is, of course, difficult to separate cause and effect. This ability may help to make a youngster more popular, or it may be that the adolescent who feels secure in his accepted status communicates more—lets himself go verbally—uninhibited by the binding restraints so often experienced by the boy who has felt the slap of rejection.

4. The specific traits stressed at each level are also of interest. The stress in the high level group upon coöperation, participation in activities, and leadership highlights the values at this economic level; the emphasis on conscientiousness and good manners at the middle is evidence of this middle group's strong need for upward mobility; the low economic level seems to put more stress on fair play and trustworthiness.

The technique of having adolescents characterize peers in their own terms seems valuable. It can be used to provide hypotheses for new avenues of research as well as to substantiate findings from other studies.

F. SUMMARY

Boys aged 13 to 15—1,050 from high income families, 775 from middle income families, and 650 from low income families—were each asked to choose up to four classmates as possible seatmates and to reject up to four others. They were asked also to list adjectives which were, in their opinion, descriptive of these accepted and rejected peers.

In many areas of personality all three income-level groups were in agreement about their standards for acceptance and rejection. Of special interest, however, is the similarity of the middle and low income levels in virtually all the areas where group standards were not in complete agreement. In such areas, these two economic levels revealed a single set of values; the high-income group had quite different standards.

All income groups agreed in characterizing accepted peers as intelligent, fair, able to take a joke, good company, athletic, quiet, conscientious, and honest; and in describing rejected peers as pesty, noisy, conceited, silly, and effeminate.

In the middle and low groups but not in the high group, community of interests, the ability to talk well, and the tendency to mind one's own business were also considered important when choosing a seatmate. In these two groups, "always fighting" was a frequent reason for rejection.

In the high-income group, but not in the middle and low groups, highly valued for acceptance (in addition to traits important to all three income groups) were coöperativeness, leadership, participation in activities, cheerfulness, and good scholarship. Leading to rejection as potential seatmates were low intelligence, lack of athletic ability, lack of leadership, non-participation in activities, and immaturity.

The results of this study should help to clarify existing concepts concerning the standards used by boys at the junior high school level in accepting or rejecting their peers.

REFERENCES

1. CROW, L. D. Teenage traits admired in members of the opposite sex and teenage interests. *National Council on Measurements Used in Education Yearbook, 1955.* (Part II, pp. 68-75.)
2. FEINBERG, M. R. Relation of Background Experience to Social Acceptance. *J. Abn. & Soc. Psychol.*, 1953, **48**, 206-214.
3. JENNINGS, H. H. Leadership and Isolation. New York: Longman-Green, 1950.
4. JONES, H. E. Development in Adolescence. New York: Appleton -Century, 1943.
5. PLATZER, K. H. Selective Processes in Junior High School: Sociometric choice Process and Personality Accomplishments. (Unpublished doctor's dissertation, New York University, 1956.)
6. TRYON, C. Evaluation of Adolescent Personality by Adolescents. *In* Barker, R. G., Kounin, J., & Wright, H. F. (*Eds.*) *Child Behavior and Development.* New York: McGraw-Hill, 1943. (Pp. 545-566.)

Baruch School of Business Administration
The City College
17 Lexington Avenue
New York 10, New York

The Journal of Genetic Psychology, 1958, **93**, 149-152.

STEREOTYPED ASSOCIATIONS TO SCHEMATIC FACES*

Department of Neurology and Psychiatry, Northwestern University Medical School

HARVEY NASH[1]

A. PROBLEM

Brunswik and Reiter (1), Samuels (4), and Halstead (2, 3) have studied the traits attributed to schematic faces. Various kinds of judges have been found to respond to schematic faces in a consistent manner. Similar ideas and feelings have been aroused by these faces in judges residing in Chicago, Boston, and Vienna. Furthermore, stereotyped associations to schematic faces have shown remarkable resistance to mental illness and to brain disease (2, 3).

In view of these consistencies, it is of interest to examine a set of data which, while providing further evidence of stereotyped associations to schematic faces, suggests that these associations are by no means universal.

B. METHOD

Forty-four boys attending a junior high school in California served as judges. They ranged in age from 11 years, 8 months to 16 years, 2 months (median age: 13 years, 7 months). The judges were presented with nine schematic faces varying in height of forehead, distance between eyes, height of chin, and length and position of nose. These nine faces were also employed by Halstead, who selected them from larger groups of faces employed by Brunswik and Reiter, and by Samuels.

Judges were presented with 14 characteristics representing polar extremes of seven variables (see Table 1), and were asked to select the face which best represents each characteristic.

For purposes of scoring, one characteristic of each pair is considered desirable, while the opposite characteristic is considered undesirable. Responses were scored as conforming or failing to conform to the stereotyped choices observed in earlier studies. Different standards were used in scoring conformity of choice for desirable and undesirable characteristics.

---

*Accepted for publication by Harold E. Jones of the Editorial Board, and received in the Editorial Office on June 19, 1957.

[1] The author, who conducted this study while attending the University of California, Berkeley, California, is indebted to H. N. McClellan and James Berry, of the Willard Junior High School, Berkeley, California, for their coöperation, and to the late Egon Brunswik, who provided encouragement and stimulus materials.

Responses of the junior high school boys are compared to the responses of 102 normal adults (males and females residing in the mid-west) reported by Halstead. The procedure employed by Halstead was followed in the present study, except that the nine faces were presented to the judges in one grouping rather than in two different groupings.

C. RESULTS

Table 1 indicates the degree to which the responses conform to the stereotyped choices reported in earlier studies. Findings for the adolescent judges are compared with those for Halstead's judges. Separate entries are made in the table for desirable (*D*) and undesirable (*U*) characteristics, since the scoring procedure results in a chance expectancy of 33.3 per cent for the former and 55.6 per cent for the latter.

TABLE 1
PERCENTAGE OF JUDGES HAVING STEREOTYPED ASSOCIATIONS

	Junior high school boys		Halstead's normal adults	
	D	U	D	U
Gay	67		86	
Sad		81		95
Beautiful	76		92	
Ugly		73		92
Good character	40		83	
Bad character		77		76
Intelligent	33		68	
Unintelligent		61		69
Likeable	52		86	
Unlikeable		87		86
Young	95		90	
Old		74		89
Energy and determination	29		60	
Lacks energy and determination		70		86

Twelve out of 14 of the percentages obtained by the adolescents are greater than would be expected by chance. Table 2 indicates that the adolescents obtained percentages which differ significantly from values expected by chance; *t*-tests were performed separately for desirable and undesirable characteristics, and the results of these two tests were combined to yield a single test of significance (6).

Despite the fact that the adolescents clearly manifest the stereotyped behavior reported by previous investigators, the percentages they obtained are exceeded by Halstead's normal adults 11 times out of 14. *T*-tests comparing the mean percentages for Halstead's adults and for the adolescents reveal that the responses of the adolescents are significantly less stereotyped than are those of the adults (see Table 2).

TABLE 2
t-TESTS RELATING TO MEAN PERCENTAGES
(All tests are 2-tailed)

Adolescents compared to chance expectancy	D	2.21	NS
	U	6.11	.001
	D and U, combined	——	.001
Halstead's adults compared to adolescents	D	4.06	.01
	U	2.68	.05
	D and U, combined	——	.001

D. DISCUSSION

Similar but by no means identical associations to schematic faces are found among judges varying in age, sex, cultural background, and intellectual soundness. The difference between the adolescent and adult judges suggests that the stereotyped associations found in adults are present, but not yet firmly established, in adolescence.[2] It would be interesting to learn whether this incomplete development of the adult stereotypes during adolescence is due to the existence of another set of stereotypes which has not yet been fully supplanted by, or altered into, the adult stereotypes.

In addition to the evidence for an over-all stereotype based on all seven response variables combined, sufficient evidence has now accumulated from several studies to permit generalizations relating to stereotypes for individual response variables. A very high consistency of agreement is found in judgments for the variables Gay–Sad and Young–Old, suggesting a widespread stereotype in attributing these qualities to the Brunswik-Reiter schematic faces. Agreement is consistently poor, on the other hand, when judgments are made for the variable Intelligent-Unintelligent. It would seem as if mood and age, but not intelligence, are commonly thought to be revealed by a limited number of spatial arrangements of the gross features of the face.

A study by Secord, Dukes, and Bevan (5) suggests that the results obtained with the use of schematic faces may not be completely applicable to more realistic stimuli. These investigators, studying impressions gained from *photographs* of faces, found striking agreement on judgments of Cheerfulness and Age, but also found good agreement on judgments of Intelligence.

---

[2] Previous findings make it seem unlikely that the difference in the stereotyped associations of the adolescent and adult judges are accounted for by differences in the sex or in the geographical origin of the judges. It is felt that the age difference accounts for the difference in the results of the two groups.

The discrepant results obtained with schematic faces and with photographs of faces might well be accounted for by analyzing the dependence of the stereotyped associations on the finer nuances of the stimulus, in the manner of Secord, Dukes, and Bevan.

REFERENCES

1. BRUNSWIK, E., & REITER, L. Eindrucks-Charaktere schematisierter Gesichter. *Z. Psychol.*, 1937, **142**, 67-134.
2. HALSTEAD, W. C. Biological intelligence. *J. Pers.*, 1951, **20**, 118-130.
3. ———. Brain and intelligence. *In* L. A. Jeffress (*Ed.*), *Cerebral Mechanisms in Behavior: The Hixon Symposium*. New York: Wiley, 1951. (Pp. 244-279.)
4. SAMUELS, M. R. Judgments of faces. *Charact. & Pers.*, 1939, **8**, 18-27.
5. SECORD, P. F., DUKES, W. F., & BEVAN, W. Personalities in faces: I. An experiment in social perceiving. *Genet. Psychol. Monog.*, 1954, **49**, 231-279.
6. WALLIS, W. A. Compounding probabilities from independent significance tests. *Econometrica*, 1942, **10**, Nos. 3 and 4.

Department of Neurology and Psychiatry
Northwestern University Medical School
303 East Chicago Avenue
Chicago 11, Illinois

The Journal of Genetic Psychology, 1958, **93**, 153-159.

A STUDY OF THE NEEDS STATUS OF ADOLESCENTS FROM BROKEN HOMES*

Department of Psychology, Ohio State University

CLAUDE J. BARTLETT AND JOHN E. HORROCKS

A. INTRODUCTION

The purpose of this study was to investigate the extent to which the psychological needs of adolescents from homes where one parent is deceased differ from the needs of adolescents from homes where both parents are living. One of the records of the White House conference (14) has noted: "The child from the broken home has been exposed to emotional strain, deprivation, or dissension rather more fully than has the child from the unbroken home." Assuming that the foregoing statement is true, the psychological needs of a child from a broken home would well present a somewhat different pattern than the psychological needs of a child from an unbroken home.

A number of previous studies (1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12) have considered the problem of the child from the broken home, but because of conflicting findings it is difficult to evaluate and compare the results of most of the studies. Unfortunately, many of them did not use a control group for comparison, nor did they attempt any extensive use of quantitative analysis of their data. Most of the studies have tended to indicate that a broken home does have an effect upon the behavior of a child, but some (1, 10) have indicated that there was no effect.

Two hypotheses were formulated for testing in the present study: (*a*) Adolescents whose parents are deceased will tend to have satisfied a need for affection to a lesser degree than do those whose parents are still living, and (*b*) Adolescents whose parents are deceased will tend to show a greater desire to form dependent relationships than do children whose parents are living.

B. SUBJECTS

The subjects in this study were drawn from a population consisting of all the children in the eighth through the twelfth grades in a mid-Ohio city of 10,000 persons.[1] Four hundred and sixty-one children were available

---

*Received in the Editorial Office on July 8, 1957.

[1] The present study is one in a series reporting the results of a community-wide study of adolescent needs, activities, and problems of emotional and social adjustment.

for study, and from this sample an experimental and a control group were drawn. Of the 461 children, the 44 who had one parent deceased were called the "experimental" group. Sixteen were females and 28 were males. From the remaining 417 children with both parents living, a random sample of 44, stratified for sex (16 females, 28 males), was drawn. Hence, the sample used in the present study consisted of 88 persons.

In order to determine the psychological needs of the subjects, the experimental form of the Horrocks-Lucas Needs Questionnaire (6) was administered.

This questionnaire consisted of two equated 90-item forms, one for boys and one for girls. The questionnaire was so constructed that the examiner can evaluate responses of the subject's perception of himself and of his perceived goals for himself. The items consisted of statements about fictitious boys or girls behaving in fictitious situations. Respondents answered each item in terms of whether or not they were like the fictitious character and whether or not they wanted to be like him. The questionnaire was scored on the basis of eight basic need factors found by Lucas (6). These factors are as follows: achievement, affection, conformity, dependence, heterosexual attraction, independence, mastery dominance, recognition.

It was felt that it might be useful to include several variables, other than the needs variables, in this study. The variables that were included are as follows: number of siblings, sex, age, number of friends, socio-economic status.

Information concerning the number of friends was obtained from the last page of the needs instrument. The subjects were asked to list all of the boys or girls (or both) who were in their favorite crowd or group. Since a subject might think he had more friends than he really did, the friends listed on the last page were not considered to be friends unless they also named the subject as a member of their groups. Thus, the number of friends, used as a variable, might be called the number of reciprocal friends from the same population as used in this study.

The socio-economic status was based on three factors which were known about the subjects: father's occupation, house type, dwelling area. These three factors were weighted to obtain a composite socio-economic score as suggested by Warner, Meeker, and Eells (13).

The criterion used to define the adolescent from a broken home was the number of parents deceased. The maximum score on this category was one, since no subjects used in this study had more than one parent deceased. A second criterion also included was the differentiation between mother or father deceased. It was not possible to carry the criterion through the entire statistical analysis since it could only be scored for the subjects who had one parent who was deceased.

D. RESULTS

An inter-correlation table of Pearson product-moment correlations was computed for the 13 variables and two criteria.

The obvious method of testing the hypotheses would be to examine the correlations between the criteria and the specific needs questionnaire categories in question. However, it should be noted that the Lucas factors, by which the questionnaire was scored, were oblique and are hence themselves inter-correlated. Also, preliminary inspection of the correlation matrix indicated that the data showed promise of yielding independent factors related to the criterion. Hence, it was decided that a factor analysis would yield more clear-cut tests of the hypotheses.

Seven factors were extracted and were rotated to maximize meaningfulness. The rotated factors are all orthogonal. The complete list of factor loadings appears in Table 1. The original correlations and the final residuals are given in Table 2.

Factor *A* might be called *Heterosexual Striving versus satisfaction from*

TABLE 1
ORTHOGONAL FACTOR LOADINGS

	A	B	C	D	E	F	G	h^2
(0)—Number of parents deceased	(23)	(29)	(36)	00	—10	—03	—04	28
(1)—Achievement	04	—14	—04	(63)	(—27)	—01	11	50
(2)—Affection	(—21)	—10	—02	(34)	05	15	(28)	27
(3)—Conformity	—15	01	02	(73)	—15	(—27)	—02	65
(4)—Dependence	—01	10	—10	(37)	(21)	(32)	—04	31
(5)—Heterosexual attraction	(87)	—17	—02	02	—01	—08	(32)	90
(6)—Independence	02	—17	16	(65)	—16	18	(—37)	67
(7)—Mastery Dominance	—08	—11	—14	(57)	—08	(25)	02	43
(8)—Recognition	(—44)	06	—03	(55)	06	—03	30	60
(9)—Number of siblings	08	(34)	(33)	—19	—17	—01	—02	30
(10)—Sex (female is higher)	—03	10	—10	—02	—17	(64)	01	46
(11)—Age	—02	+12	+04	00	(—76)	08	—08	61
(12)—Number of friends	—05	(—62)	03	08	—02	(47)	07	62
(13—Socio-economic status	—01	(87)	03	—02	—10	00	09	78

( ) Indicates significance.

TABLE 2

Intercorrelations of Variables and Criteria and Final Residuals

	0	0′	1	2	3	4	5	6	7	8	9	10	11	12	13
(0) Number of parents deceased	(xx)	00	00	—15	—02	—05	12	04	—06	—14	26	00	20	—13	26
(0)′ Mother or father deceased (mother is higher)*	—	(xx)	—13	00	—24	02	04	—17	—12	01	34	19	12	—04	27
(1) Achievement	03	—	(xx)	30	51	21	14	43	43	42	—10	02	14	22	—01
(2) Affection	—04	—	06	(xx)	23	16	—07	11	32	32	—10	—09	—03	17	—11
(3) Conformity	—02	—	02	01	(xx)	13	—15	48	46	42	—17	—07	+09	—11	—02
(4) Dependence	—01	—	05	—01	—02	(xx)	—08	27	31	24	—15	24	—08	04	05
(5) Heterosexual attraction	—02	—	03	01	—05	—02	(xx)	—04	—03	—31	—02	—03	—01	00	—13
(6) Independence	00	—	00	—04	03	02	04	(xx)	50	23	—19	11	22	14	20
(7) Mastery-Dominance	04	—	03	06	09	03	02	08	(xx)	37	—25	16	05	29	—17
(8) Recognition	—02	—	09	—04	—04	03	—03	02	03	(xx)	—21	—03	—01	—03	05
(9) Number of siblings	—01	—	03	04	—06	—04	—02	—08	—06	—07	(xx)	03	10	—15	38
(10) Sex (female is higher)	02	—	00	01	09	06	05	02	—01	—01	00	(xx)	13	25	07
(11) Age	+08	—	—04	+02	—01	+03	+05	+07	—01	+05	—08	—06	(xx)	—07	17
(12) Number of friends	06	—	08	—02	—04	07	—05	—08	06	—06	07	01	—05	(xx)	—51
(13) Socio-economic status (lower is higher)	00	—	09	—03	—03	00	—01	—02	—06	—01	07	—03	00	02	(xx)

*Omitted factor analysis.

Parents. This factor might be best interpreted as an indication that the adolescents who have one parent deceased are neither receiving as much affection at home, nor as much recognition from adults and peers as are the adolescents from normal homes. In order to compensate for this lack of affection and recognition, they seem to be striving for attention from the opposite sex.

Because of the high loading of the socio-economic status on Factor *B*, it might be considered a *Socio-economic* factor. This factor can be interpreted as an indication that families which have many children and in which one of the parents is deceased (most likely the father) tend to be of a lower socio-economic status. The adolescents who come from these homes where one parent is deceased, where there is a large number of siblings, and where they are of low socio-economic status, show a tendency to have fewer friends.

The only other factor with a significant loading on the criterion is Factor *C*. The name which seems appropriate for this factor is *Death of Mother as Related to Childbirth.* It appears that the more children there are in the family, the more probable it is that one of the parents is deceased. It probably arises from the hazards of childbirth. The only correlation with the second criterion, mother or father deceased, which was statistically significant (at the .05 level) was the correlation between number of siblings and mother deceased. For this reason, the factor probably refers to death of the mother rather than death of one parent.

Factor *D* has significant loadings on the seven needs variables that show degree of need satisfaction. The heterosexual attraction variable shows striving to satisfy a need rather than need fulfillment. Thus, an appropriate name for this factor seems to be *Degree of Need Satisfaction.*

Age is the most heavily weighted variable on Factor *E*. The factor might be called *Age as Related to Differential Need Satisfaction.* It might be interpreted as a tendency for younger adolescents to become more dependent upon their parents and friends than are older children. The negative loading on achievement indicates that the younger adolescents either achieve less or are less interested in achievement.

Factor *F* might be called *Sex as Related to Friends and Differential Need Satisfaction.* Female adolescents show a tendency to have more friends than do males. They also show a greater desire to form dependent relationship with their friends or family. These girls show a tendency not to conform to adult standards, and they also appear to be trying to dominate their friends and family. One might speculate that these girls are forming dependent relations as a means of dominating their friends and families. Also, the fact that they tend to have more friends may tend to increase their dependence upon those friends.

Factor *G* indicates that the adolescents of this study do not seem to be independent from adults and appear to be satisfying the need of affection and recognition from adults and peers. This recognition is probably from peers of the same sex, since the adolescents of the present study show a tendency of striving for rather than satisfying the need for heterosexual attraction. This factor might be named *Heterosexual versus Adults and Peers of Same Sex as Source of Recognition.*

E. Summary and Conclusion

The purpose of the present study was to determine how the needs status of adolescents from homes where one parent is deceased differs from the needs status of adolescents from homes where both parents are living. The Experimental Form of the Horrocks-Lucas Needs Questionnaire was used to determine needs status. Since it is probable that other factors may also influence the needs status, some of the more important variables were included in the study of controls.

Inter-correlations of all the variables and criteria were computed. Then, the data were factor-analyzed. The following seven factors were extracted:

Factor *A*—Heterosexual Striving versus Satisfaction from Parents.
Factor *B*—Socio-economic Factor.
Factor *C*—Death of Mother as Related to Childbirth.
Factor *D*—Degree of Need Satisfaction.
Factor *E*—Age as Related to Differential Need Satisfaction.
Factor *F*—Sex as related to Friends and Differential Need Satisfaction.
Factor *G*—Heterosexual versus Adults and Peers of Same Sex as Source of Recognition.

In only three factors (*A*, *B*, and *C*) was the loading on the criterion significant. Of these three, ony one (*A*) showed any significant relationship between needs status and number of parents deceased. This factor indicates that the adolescents from homes where one parent is deceased tend to receive less recognition and affection from adults. In order to compensate for this lack of recognition and affection, they seem to be striving for attention from the opposite sex. This factor tends to support one of the hypotheses made at the beginning of the study: "The group whose parents are deceased will tend to have satisfied the need for affection to a lesser degree than the control group." No conclusive evidence was found to support or reject the other hypothesis made, that: "The group whose parents are deceased will tend to show a greater desire to form dependent relationships than the control group." If the desire for attention from the opposite sex was interpreted as also being a desire to form dependent relationships, one might find some support for this second hypothesis.

The major hypothesis of the present study was that the loss of the father or mother creates a change in the home situation that can affect the psychological needs of a child in such a home. This study appears to offer some evidence to support this hypothesis.

REFERENCES

1. ADAM, J. An inquiry into the influence of broken homes on the maladjustment of children. *Brit. J. Educ. Psychol.*, 1946, **16**, 45.
2. BEALS, L. M. Study of certain home factors and their relationships to the personal adjustment of children. *Sch. & Soc.*, 1950, **72**, 55-57.
3. BEHRENS, M. L. Child rearing and character structure of the mother. *Child Development*, 1954, **25**, 225-238.
4. CAMPBELL, M. W. Effect of the broken home upon the child in school. *J. Educ. Sociol.*, 1932, **5**, 274-281.
5. DUVALL, E. M., & NEISSER, E. G. Doing without dad. *Nat. Parent-Teach.*, 1944, **38**, 7-9.
6. LUCAS, C. L. An emergent category approach to the analysis of adolescent needs. (Unpublished doctoral dissertation, The Ohio State University, 1951.)
7. MAREON, M. The broken home as a factor in the behavior of fifty-six children. *Smith Coll. Stud. Soc. Work*, 1931, **2**, 80.
8. MEYERS, G. C. How home and family conditions affect the child's personality. *Nat. Elem. Princ.*, 1936, **15**, 303-310.
9. ROUMAN, J. School child's problem as related to parental factors. *Understand. Child*, 1955, **24**, 50-55.
10. SLVERMAN, B. The behavior of children from broken homes. *Amer. J. Orthopsychiat.*, 1935, **5**, 11-18.
11. TORRANCE, P. Influence of the broken home on adolescent adjustment. *J. Educ. Sociol.*, 1945, **18**, 359-364.
12. WALLENSTEIN, N. Character and personality of children from broken homes. *Teach. Coll. Rec.*, 1938, **39**, 747-748.
13. WARNER, W. L., MEEKER, M., & EELLS, K. Social Class in America. Chicago: Science Research Associates, 1949.
14. WHITE HOUSE CONFERENCE ON CHILD HEALTH AND PROTECTION. *The Adolescent in the Family.* New York: Appleton-Century-Crofts, 1934.

Department of Psychology
The Ohio State University
Columbus 10, Ohio

The Journal of Genetic Psychology, 1958, **93**, 161-163.

BOOKS

Now that there is a special *APA* journal completely devoted to the publication of book reviews, it is no longer necessary that other journals emphasize such publication. It has always been our conviction that book reviews are a secondary order of publication unless they carry information that is equally as important as the book. However, the publication of book titles is a very important service, and we shall continue to render that service.

In any given issue of this journal, we may continue to publish one or more book reviews, but we do not consider such publication a major function of this journal. In line with this policy, we can no longer pay for such manuscripts.

BOOKS RECENTLY PUBLISHED

AMBROSE, G., & NEWBOLD, G. Hypnosis in Health and Sickness. London: Staples, 1958. Pp. 196.

BATESON, G. Naven. Stanford: Stanford Univ. Press, 1958. Pp. 312.

ELLIS, A. Sex Without Guilt. New York: Lyle Stuart, 1958. Pp. 200.

EX, J. Situatie-Analyse en Sociaal-Psychologisch Experiment. Bussum: Uitgeverij Paul Brand, 1957. Pp. 130.

FARNSWORTH, P. R. The Social Psychology of Music. New York: Dryden, 1958. Pp. 304.

FEIGL, H., SCRIVEN, M., & MAXWELL, G., *Eds.* Minnesota Studies in the Philosophy of Science. Minneapolis: Univ. Minnesota Press, 1958. Pp. 553.

GROSS, N., MASON, W. C., & MCEACHERN, A. W. Explorations in Rôle Analysis. New York: Wiley, 1958. Pp. 379.

HADLEY, J. M. Clinical and Counseling Psychology. New York: Knopf, 1958. Pp. 682.

HERZBERG, F., MAUSNER, B., PETERSON, R. O., & CAPWELL, D. F. Job Attitudes. Pittsburgh: Psychological Service of Pittsburgh, 1957. Pp. 279.

HOLLINGSHEAD, A. B., & REDLICH, F. C. Social Class and Mental Illness. New York: Wiley, 1958. Pp. 441.

HUTT, M. L., & GIBBY, R. G. The Mentally Retarded Child. Boston: Allyn & Bacon, 1958. Pp. 334.

JOURARD, S. M. Personal Adjustment. New York: Macmillan, 1958. Pp. 462.

LANDRETH, C. The Psychology of Early Childhood. New York: Knopf, 1958. Pp. 412.

LIN, H. The Moon Vow. New York: Pageant, 1958. Pp. 312.

MEADOWS, T. Psychology of Learning and Teaching Christian Education. New York: Pageant, 1958. Pp. 393.

MERRY, F. K., & MERRY, R. V. The First Two Decades of Life. New York: Harper, 1958. Pp. 642.

MOSCROP, M. In-Service Training for Social Agency Practice. Toronto: Univ. Toronto Press, 1958. Pp. 237.

O'NEIL, W. M. An Introduction to Method in Psychology. New York: Cambridge Univ. Press, 1957. Pp. 155.

PIDDINGTON, R. An Introduction to Social Anthropology. London: Oliver & Boyd, 1957. Pp. 819.

PIKUNAS, J. Fundamental Child Psychology. Milwaukee: Bruce, 1957. Pp. 259.

PRESCOTT, D. A. The Child in the Educative Process. New York: McGraw-Hill, 1957. Pp. 502.

REMMERS, H. H., & RADLER, D. H. The American Teenager. Indianapolis: Bobbs-Merrill, 1957. Pp. 267.

Rifkin, A. H., *Ed.* Schizophrenia in Psychoanalytic Office Practice. New York: Grune & Stratton, 1957. Pp. 150.

Rogers, D. Mental Hygiene in Elementary Education. Boston: Houghton-Mifflin, 1957. Pp. 497.

Roman, M. Reaching Delinquents Through Reading. Springfield: Thomas, 1957. Pp. 125.

Rittwagen, M. Sins of Their Fathers. Boston: Houghton-Mifflin, 1958. Pp. 264.

Rothstein, J. Communication, Organization, and Science. Indian Hills, Colo.: Falcon's Wing Press, 1957. Pp. 110.

Sanders, I. T. The Community. New York: Ronald Press, 1958. Pp. 431.

Schramm, W. Responsibility in Mass Communication. New York: Harper, 1957. Pp. 391.

Schreier, F. T. Human Motivation. Glencoe: Free Press, 1957. Pp. 256.

Schwartz, A., & Tiedeman, S. G. Evaluating Student Progress in the Secondary School. New York: Longmans, Green, 1957. Pp. 434.

Scott, J. P. Animal Behavior. Chicago: Univ. Chicago Press, 1958. Pp. 281.

Scott-Maxwell, F. Women and Sometimes Men. New York: Knopf, 1957. Pp. 207.

Sears, R. R., Maccoby, E. E., & Levin, H. Patterns of Child Rearing. Evanston: Row, Peterson, 1957. Pp. 549.

Service, E. R. A Profile of Primitive Culture. New York: Harper, 1958. Pp. 474.

Shneidman, E. S., & Farberow, N. L., *Eds.* Clues to Suicide. New York: McGraw-Hill, 1957. Pp. 227.

Simon, B., *Ed.* Psychology in the Soviet Union. Stanford: Stanford Univ. Press, 1957. Pp. 305.

Skinner, B. F. Verbal Behavior. New York: Appleton-Century Crofts, 1957. Pp. 478.

Smith, K. U., & Smith, W. M. The Behavior of Man. New York: Henry Holt, 1958. Pp. 516.

Steckle, L. G. Problems of Human Adjustment. New York: Harper, 1957. Pp. 419.

Stern, E. M. Mental Illness: A Guide For the Family. New York: Harper, 1957. Pp. 95.

Stone, L. J., & Church, J. Childhood and Adolescence. New York: Random House, 1957. Pp. 456.

Stoops, E., & Wahlquist, G. L. Principles and Practices in Guidance. New York: McGraw-Hill, 1958. Pp. 369.

Stott, D. H. Unsettled Children and Their Families. New York: Philosophical, 1956. Pp. 240.

Strachey, A. The Unconscious Motives of War. New York: International Univ. Press, 1957. Pp. 283.

Strang, R. The Adolescent Views Himself. New York: McGraw-Hill, 1957. Pp. 581.

Super, D. E. The Psychology of Careers. New York: Harper, 1957. Pp. 362.

Symonds, P. M. Dynamics of Psychotherapy. New York: Grune & Stratton, 1957. Pp. 175-398.

Szasz, T. S. Pain and Pleasure. New York: Basic Books, 1957. Pp. 301.

Tashman, H. F. Today's Neurotic Family. New York: New York Univ. Press, 1957. Pp. 214.

Taylor, R. W., *Ed.* Life, Language, Law. Yellow Springs: Antioch, 1957. Pp. 223.

Templin, M. C. Certain Language Skills in Children. Minneapolis: Univ. Minnesota Press, 1957. Pp. 183.

Travis, L. E., *Ed.* Handbook of Speech Pathology. New York: Appleton-Century-Crofts, 1957. Pp. 1088.

Traxler, A. E. Techniques of Guidance. New York: Harper, 1957. Pp. 374.

Turner, M. E. The Child Within the Group. Stanford: Stanford Univ. Press, 1957. Pp. 93.

Turner, R. H., & Killian, L. M. Collective Behavior. Englewood Cliffs, N. J.: Prentice-Hall, 1957. Pp. 547.

Van de Geer, J. P. A Psychological Study of Problem Solving. Haarlem: Uitgeverij de Toorts, 1957. Pp. 214.

Vernon, M. D. Backwardness in Reading. New York: Cambridge Univ. Press, 1957. Pp. 228.

Voeks, V. On Becoming an Educated Person. Philadelphia: Saunders, 1957. Pp. 147.

Von Fieandt, K. Varseblivningens Psykologi. Helsingfors: Soderstrom, 1956. Pp. 370.

Von Mering, O., & King, S. H. Remotivating the Mental Patient. New York: Russell Sage, 1957. Pp. 216.

Wallerstein, R. S. Hospital Treatment of Alcoholism. New York: Basic Books, 1957. Pp. 212.

Walker, C. R. Toward the Automatic Factory. New Haven: Yale Univ. Press, 1957. Pp. 232.

Werner, H. Comparative Psychology of Mental Development. New York: International Univ. Press, 1957. Pp. 564.

Werre, P. F. The Relationships between Electroencephalographic and Psychological Data in Normal Adults. Leiden: Universitaire pers Leiden, 1957. Pp. 152.

Wilensky, H. L. Intellectuals in Labor Unions. Glencoe: Free Press, 1956. Pp. 336.

Wilensky, H. L., & Lebeaux, C. N. Industrial Society and Social Welfare. New York: Russell Sage, 1958. Pp. 401.

Winnicott, D. W. Mother and Child. New York: Basic Books, 1957. Pp. 210.

Wolfenstein, M. Disaster. Glencoe: Free Press, 1957. Pp. 231.

Woolf, M. D., & Woolf, J. A. Remedial Reading. New York: McGraw-Hill, 1957. Pp. 424.

Wormhoudt, A. Hamlet's Mouse Trap. New York: Philosophical, 1956. Pp. 221.

Yates, D. H. Psychology You Can Use. New York: Crowell, 1957. Pp. 248.

Zaehner, R. C. Mysticism Sacred and Profane. New York: Oxford Univ. Press, 1957. Pp. 256.

Bergler, E. The Psychology of Gambling. New York: Hill & Wang, 1957. Pp. 244.

Blackston, S. The Three Talents. New York: Greenwich Book, 1957. Pp. 93.

Brozek, J., *Ed.* Symposium on Nutrition and Behavior. New York: National Vitamin Found., 1957. Pp. 124.

Bruner, J. S., Brunswk, E., Festinger, L., Heider, F., Muenzinger, K. F., Osgood, C. E., & Rapaport, D. Contemporary Approaches to Cognition. Cambridge: Harvard Univ. Press, 1957. Pp. 210.

Carrington, W. L. Psychology, Religion, and Human Need. Great Neck, N. Y.: Channel Press, 1957. Pp. 315.

Cattell, R. B. Personality and Motivation Structure and Measurement. Yonkers-on-Hudson: World Book, 1957. Pp. 948.

Cleugh, M. F. The Slow Learner. New York: Philosophical, 1957. Pp. 186.

Corsini, R. J. Methods of Group Psychotherapy. New York: McGraw-Hill, 1957. Pp. 251.

Duvall, E. M. Family Development. Philadelphia: Lippincott, 1957. Pp. 533.

Ellis, A. How to Live With a Neurotic. New York: Crown, 1957. Pp. 206.

Ewing, A. W. G., *Ed.* Educational Guidance and the Deaf Child. Washington, D. C.: Volta Bureau, 1957. Pp. 345.

Farnsworth, D. L. Mental Health in College and University. Cambridge: Harvard Univ. Press, 1957. Pp. 244.

Festinger, L. A. Theory of Cognitive Dissonance. Evanston: Row, Peterson, 1957. Pp. 291.

Ferster, C. B., & Skinner, B. F. Schedules of Reinforcement. New York: Appleton-Century-Crofts, 1957. Pp. 740.

Fox, J. The Chronically Ill. New York: Philosophical, 1957. Pp. 229.

Gordon, C. W. The Social System of the High School. Glencoe: Free Press, 1957. Pp. 184.

$14.00 per year
$10.00 per volume
Single Numbers $5.00

QUARTERLY
Two volumes per year

December, 1958
Volume 93, Second Half

Founded by G. Stanley Hall in 1891, reorganized by Carl Murchison in 1925

THE JOURNAL OF GENETIC PSYCHOLOGY

Child Behavior, Animal Behavior, and Comparative Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a liberal contract, and may not be resold for a period of five years from the date of publication.

DECEMBER, 1958

(Manuscripts are printed in the order of final acceptance)

Affective responses to the concept of death in a population of children and early adolescents 167
By Irving E. Alexander and Arthur M. Adlerstein

Conditioned responses in the opossum 179
By W. T. James

Kinesthetic cues in the development of time concepts 185
By Sanford Goldstone, William K. Boardman, and William T. Lhamon

Conceptual block-sorting as a function of type and degree of mastery of discriminative verbal responses 191
By Albert E. Goss and Marie C. Moylan

Raven's progressive matrices: A review and critical evaluation . . . 199
By Henry R. Burke

(OVER)

Copyright, 1958, by The Journal Press
Provincetown, Massachusetts, U. S. A.

Entered as second-class matter May 11, 1937, at the post-office at Provincetown, Mass., under the Act of March 3, 1879

The development of paw preferences in cats and monkeys 229
By J. M. Warren

Reevaluation of the Revised Stanford-Binet Scale, Form *L* Vocabulary as a test of intelligence for the kindergarten and primary school child . . 237
By Boris M. Levinson

The ecological veracity of the self-report 249
By Samuel R. Pinneau and Alexander Milton

Cultural pressure and *WAIS* scatter in a traditional Jewish setting . . . 277
By Boris M. Levinson

Preirradiation fatigue as a factor in the prevention of irradiation deaths in rats 287
By W. Lynn Brown and Robert K. White

A study of seventh grade children's reading of comic books as related to certain other variables 291
By W. Paul Blakely

SHORT ARTICLES AND NOTES 303

Substitution with interrupted tasks of differing valence (Anitra Gordon and Willard Thurlow) 303

BOOKS 307

BOOKS RECENTLY PUBLISHED 307

Packard, V. *The Hidden Persuaders.* (Reviewed by Henry Winthrop) 309

The Journal of Genetic Psychology, 1958, **93**, 167-177.

AFFECTIVE RESPONSES TO THE CONCEPT OF DEATH IN A POPULATION OF CHILDREN AND EARLY ADOLESCENTS*[1]

Department of Psychology, Princeton University, and the Eastern Pennsylvania Psychiatric Institute

IRVING E. ALEXANDER AND ARTHUR M. ADLERSTEIN[2]

A. INTRODUCTION

The child's concern with death has been a part of most parents' experience. Several psychologists, in recording various aspects of development in children, including their own, have discussed this topic briefly (8, 12, 13, 17). More controlled observations on the subject of the child and death are very scarce in the psychological literature both in this country and abroad.

Perhaps the most comprehensive piece of work was done by Sylvia Anthony. In an excellent and much neglected book (2) she summarizes a series of investigations carried out in pre-war England over a two-year period. Children between 3 and 13 years of age were studied with a variety of techniques, including parents' written accounts of their children's spontaneous interest in death, a story-completion test, and an intelligence test in which material relevant to death appears.

Anthony found that death thoughts are frequent in children's fantasies. Roughly 50 per cent of the children tested on the story completion task made reference to death in completing their stories although the concept did not appear in the story stem. Five stages in children's thoughts about death were distinguished. These ranged from ignorance of the meaning of the word to clear definition in logical or biologically essential terms. By the fifth or sixth year there is meaning attached to the word but rarely in terms of the latter criteria. By the eighth or ninth year causal-logical explanation is being used to some degree by almost all children. These findings parallel those of Piaget on the development of the concept of causality.

Some years previous to this work Schilder and Wechsler (15) published the results of a study on the attitudes of children towards death. Their

---

*Received in the Editorial Office on July 31, 1957.

[1] This research was supported by a Public Health Service Grant M-1334.

[2] Our thanks go to the Grand Street Settlement Camps for permission to carry out this work and especially to Miss Dora Tannenbaum, Director, for her understanding and coöperation.

techniques included observations of play, spontaneous stories given to pictures, and direct questioning. The population observed consisted of 76 children between the ages of 5 and 15. Diagnostic categories included organic hyperkinetics, epileptics, mental defectives, behavior problems, and schizoids. The findings were not evaluated quantitatively but rather were given in general statements based on the observations of the younger and approximately normal children.

These observers concluded that children deal with death realistically—in a matter-of-fact way. They do not want to die, yet they rarely express fear. They are ready to believe that others can die but do not believe their own deaths probable. The child connects death with deprivation and aggression. These latter concepts are the substructure upon which actual experience with death and dying is built. The authors offer this framework in opposition to the philosophical speculation of Heidegger that death and time are intimately related. For the child this relationship does not seem essential.

Nagy's study (10), completed in the late 1930's in Budapest, has recently appeared in the American literature. The written compositions, drawings, and recorded discussions of children between 3 and 10 years of age were analyzed. The aim of this work was to investigate the child's *theories* concerning the nature of death. Replies to the question "What is death?" were categorized into three stages. In children under five death is reversible and not final. The youngest children are likely to interpret death as sleep or departure while the five-year-old modifies this distinction by making death gradual and temporary. In the second stage, from the fifth through the ninth year, death is most often personified. It is also thought of as an aggressive event contingent upon the actions of others. The third stage, which follows, brings with it the recognition that death is a process dependent on natural laws, one which is characterized by the cessation of vital bodily activities.

Nagy's findings are similar to those of Cousinet (6), who also distinguished three stages in children's thinking about death: (*a*) the refusal to accept the idea, (*b*) the substitution of severe but curable illness for death, and (*c*) the disappearance of death as a troublesome concept. According to this author the last stage may precede an understanding of the concept in naturalistic terms.

Several writers have been concerned with the *fear of death* in children and its origin. Anthony (2) relates it to the fear of retaliation (talionic law)—a fear of the aggression of others. In her study two kinds of anxiety about death were differentiated: chronic and critical. The first refers to a reaction from one's own aggressive impulses and is not dependent on a clear conception of the idea of death. Such a state is found in children under five. Critical anxiety is a later phenomenon appearing when the child recognizes the fact that he is an independent being and consequently can die. Schilder and Wechsler (15) state that the fear of death is rare. The child is not concerned with dying but rather with being murdered. Chadwick (5) questions the orthodox psychoanalytic belief that the fear of death is a derivation of castration anxiety and offers the possibility that the converse may be true. Infantile separation anxiety is given as a root of death fears which serve the purpose of masking the basic wish to return to the mother. Other relationships are drawn between fear of death and conditions of physical restraint, masturbation guilt, and fear of the dark.

Harnik (7) postulates that the fear of suffocation arising from difficulty in breathing in early infancy is a possible basis for the fear of death. Caprio (4) traces neurotic death fears to early fears of the dark, intense emotional experiences at funerals, and the influence of adults with morbid superstitions about death. Rosenzweig and Bray (14) point out an interesting fact for the hypothesis that fear of death relates to extreme guilt about aggressive death wishes towards others. They found that schizophrenics experienced sibling death more often than normals in the population they studied.

Other ideas on the fear of death appear in sources not easily available. Kotsovsky (9) designates several influencing factors: fear of pain, fear of the unknown, the external unpleasant experiences associated with death and burial, and the disappointment and guilt concerning unfulfilled potential. Kotsovsky speaks of both a conscious and an unconscious fear of death. Osipov (11) calls intense fear of death a neurological symptom. It is a product of the transitional stage between the naiveté of the child and the maturity of the adult.

There seems to be reasonable agreement in these studies that the child's thinking about death progresses from a state of non-awareness of the meaning of the concept through a series of intermediate steps to a point where death is considered in logical, causal, naturalistic terms. The degree of affective involvement that children have with death is not quite so clear. On the one hand there are the statements of indifference shown by children: matter-of-factness, reality orientation, and lack of fear. At the same time there are reports of denial (in the defensive sense), morbid concern, and undue stress in fantasy production. With regard to affective arousal, the indicators used by the above investigators were gross. They were mainly clinical judgments of play, fantasy, and discussion. The data did not lend themselves easily to quantification or to reliability or validity checks. How do children of various ages respond affectively to the idea of death? This question was the focal point of our investigation. It was our intent to approach this question with quantitative methods.

B. METHOD

1. *Technique and Rationale*

The essentials of the method have been reported previously (1). It consisted of a word association task to which response times, galvanic skin responses (*GSR*), and response words were recorded.

The stimulus words, 27 in all, were taken from a manual for the teaching of spelling published by the Board of Education of the City of New York (19). The word groupings in this manual were based on word frequency counts of written English for elementary and high school children. Our final list was composed of three sets of words. Each set was equivalent to every other in terms of mean frequency of usage in the language, mean length. and number of syllables. One set was designated as "a priori affective," that is, words that would by general agreement be likely to arouse affect in children. These words were drawn from two general concepts, sex and family. There were six such words in all: mama, papa, child, love, married, and kiss. A second set was designated as "basal" words, that is, words with no apparent general relationship to each other or to other words on the list. These are samples from the language against which to measure our critical words. Examples of basal words are dress, brave, happy, star, deep, speak, animal, penny, flat, etc. The third set, and critical to this experiment, were words related to one another under the superordinate heading of "death." These were buried, kill, and dead. They appeared in positions 6, 15, and 24 on one list and were preceded and followed by two basal words.

The rationale is as follows: the difference between response to a priori affectives and basal words gives us a test of the validity of our measuring technique. The crucial test is between responses to death words and responses to basals. It is to be emphasized that basal words were by no means chosen to be affect-free as inspection will reveal. They are samples from the language, each of which can have affective meaning for any subject. Our concern is whether death words are responded to with greater latencies and increased *GSR*s than are basal words.

2. *Subjects*

The subjects used in this study were 108 males between the ages of 5 and 16. The age distribution is given in Table 1.

The children were all summer campers at the Grand Street Settlement Camps, Craig's Meadow, East Stroudsburg, Pennsylvania. For the most part they were from low income families. Both negroes (12 per cent) and whites (88 per cent) were represented in the group. None of the children were known to have psychiatric histories nor was there any reason to believe that intelligence was distributed atypically in any of the age groupings. Aside from trying to get a representation of children at all available ages, the only other selection criterion was willingness to serve as a subject. The selection procedure was as follows: an entire group would be scheduled for testing during one of the activity hours in a special room not used by the campers for any other purpose. The experimenters spoke to the group about the alleged purpose of the experiment—to see whether a recording machine (the Sanborn Twin-Viso Recorder Model 651-42) was adequate for measuring changes that went on inside the body. Comparisons with the stethoscope, *EKG*, etc., were made depending upon the experience level of the children. Following this preliminary explanation and discussion a group demonstration was held. The group counselor was used as a subject and the children observed how the Sanborn recorded and how the time record was made. In all groups great enthusiasm was displayed for being given the opportunity to be "guinea pigs." The demonstration served to allay undue anxiety about the nature of the procedure. After the demonstration volunteers were solicited and the day's testing schedule set up. The volunteer rate was roughly 95 per cent.

TABLE 1
AGE-FREQUENCY DISTRIBUTION OF SUBJECTS

Age	Number
5	1
6	5
7	10
8	13
9	10
10	17
11	12
12	9
13	10
14	10
15	10
16	1

3. *Procedure*

In the individual testing the subject was seated at a table opposite the experimenter. His right hand, both palm and dorsal surface, was rubbed vigorously with electrode jelly to reduce basal skin resistance (3). The hand

Bureau Ednl. Psy. Research
DAV'D ING COLLEGE
Dated
Acc. No

was then placed in a foam rubber rig that encased two zinc plated electrodes, one attached to a wooden board secured to the table, the other to an adjustable webbed strap secured to the board. Contact between the hand surfaces and the electrodes was mediated by two foam rubber inserts dipped in a salt water solution. The advantage of this particular rig was to immobilize the hand without discomfort. A Western Electric telephone receiver housed in a sound-proofed wooden box (6″ x 14″ x 6″) was then placed close to the mouth of the subject. He was reminded to respond into the telephone receiver with the first word that came to mind when a stimulus word was presented. During this time the subject's basal resistance was gauged and balanced by use of a Wheatstone bridge technique. Four balance levels were used: 5,000, 10,000, 15,000, or 20,000 ohms. Roughly 90 per cent of the subjects performed at a basal resistance level of 10,000 ohms or less. Resistance was added in series to that of the subject to bring him up to the closest balance level. This procedure allowed for the measurement of change in skin resistance from a known basal level. In the analysis of the data all responses were transformed by appropriate constants to units relative to a 5,000-ohm basal resistance.

The telephone receiver was employed as a voice key to interrupt a Standard electric precision timer (Model S1). This instrument, accurately read to the nearest one hundredth of a second, was initiated manually by the experimenter upon presentation of a stimulus word. One experimenter recorded response times while the other monitored the continuous recording of the change in skin resistance. A practice word list leading into the regular list was given to settle the subjects in the apparatus. The time interval between word presentations varied between 20 seconds and one minute, depending on the rapidity of the subject's return to a resting level after responding.

C. RESULTS

For purposes of analysis our subject population was divided into three subgroupings: age 5 through 8 ($N = 29$), 9 through 12 ($N = 48$), and 13 through 16 ($N = 31$). This particular division was prompted by several considerations. It seems to be one that violates least the observations that have been made about the development of logical thought in children. Piaget (12) places the beginning of this event between the seventh and eighth year and its consistent utilization in everyday action at age 12. Anthony (2) says that between 5 and 6 children first begin to have some understanding of death. However, it is not until 8 or 9 that reference is made to logical and biological essentials in discussing the concept. Nagy (10) also distinguished the period between 5 and 9 as being a unique one in which death is usually personified. This division also corresponds roughly to parts of Sullivan's (16) developmental scheme, the stage of peer group importance (5-8), isophilic love (9-12), and the development of the lust dynamism (13-). The break at 13 is in line with physical notions about adolescence, in particular the maturing of secondary sex characteristics.

Response data originally in millisecond and millimeter units were skewed. A logarithmic transformation served to normalize the distributions. As a rough check on the validity of the technique for measuring affective arousal, difference *t*'s between responses to a priori affective words and basal words for all subjects were assessed. For both indices these differences were significant ($p_{gsr} < .01$, $p_{time} < .05$). We may therefore dispense with the a priori affective category from further consideration. The remaining data were analyzed by the analysis of variance technique. Table 2 presents the results.[3, 4, 5]

TABLE 2
ANALYSIS OF VARIANCE OF *GSR* AND RESPONSE TIME SCORES

	GSR $N = 88$				Response Time $N = 108$			
Source	Sum of squares	*df*	Mean square	*F*	Sum of squares	*df*	Mean square	*F*
Treatments	.41	1	.41	4.1**	1.62	1	1.62	14.73**
Age Groups	2.69	2	1.35	.54	20.48	2	10.24	15.28**
Treatments x Age Groups	1.74	2	.87	8.7**	.21	2	.11	1.00
Subjects	213.35	85	2.51		69.93	105	.67	
Treatments x subjects	8.85	85	.10		12.01	105	.11	

$^{**}p < .01$.

From this table it is clear that the treatment variable is significant on both response indicators. Death words are responded to by the total group with increased latency and decreased skin resistance. On the *GSR* measure the several age groups respond differently to the treatment variable, while on the response time measure the several age groups respond differently to words as a whole independent of treatment. This latter statement reflects the

---

[3] The treatments *x* subjects variance is used as the error term for treatments *x* age groups. The subjects variance is used as the error term for age groups. See: Wilk and Kempthorne (18).

[4] The *N*'s differ for the *GSR* and response time measures. In 20 cases *GSR* records were either not obtainable or not scorable. No one sub-group contained a preponderance of these people.

[5] Individual subject means for basal and death words were based on different numbers of observations (18 and 3). This difference was handled by a weighting technique in the analysis.

common observation that response time in a word association task decreases as a function of age (at least for childhood and early adolescent years). These results indicate that death is associated with an increase in affect in our total population. Now let us see how this phenomenon operates developmentally. Table 3 presents difference t's (basal vs. death) for the age sub-divisions.[6]

TABLE 3
t VALUES FOR WORD GROUP COMPARISONS BY AGE GROUPS ON BOTH RESPONSE MEASURES

Age group	*GSR*	Response time
5- 8	1.65*	4.74**
9-12	.10	3.17**
13-16	3.47**	2.80**

*p .05.
**p .01.

This analysis discloses that in all age groups children respond with greater latencies to death words than to experimentally equivalent basal words. For skin resistance changes the picture differs. In the youngest age group there is a significant decrease ($p < .05$) in skin resistance to death words and this same change is also in evidence in the adolescent group ($p < .01$). The 9 through 12 group, however, shows no reliable difference.

D. DISCUSSION

In this population death words generally elicit indications of increased emotional response. This increase we shall attribute to both individual and cultural factors, each of which is represented differentially in our measures. There are, however, wide individual differences in response both among children and among age groups.

Of the two response indicators, the time measure is probably much more sensitive to cultural influence and this is why all groups that we have tested show longer response time to death words than to basal words. That is to say, death is not welcomed as a conversation piece in our culture regardless of one's attitudes towards the subject. It would seem that increased latency could either reflect one's own distaste for the subject or the awareness of the impact of this topic upon others. In any case the combined possibilities could account for the fact that these latency differences are found throughout the age range tested.

The *GSR* measure, we feel, is more sensitive to the inner feeling states of the individual, to the immediate emotional value that the stimulus has for him. One would then have to turn to both theory and observation in developmental psychology to account for the differences in response among the several age groups on this measure. What we must explain is the similarity in response for the 5 through 8 and 13 through 16 groups, and the difference between these two and the 9- through 12-year group. The most general statement we would make is that death has a greater emotional significance for people with less stable ego self-pictures than for people with an adequate concept of the self. The period from 5 through 8 has been pointed out as a trying one by many theorists. In the Freudian scheme it is the time of the resolution of the Oedipal situation, and the consequent search for a more permanent and acceptable identification figure. The superego is born and with it comes the attendant grief for the ego in learning how to cope with this harsh master. For Sullivan it is the stage in which peers become important, when new authority figures emerge, when the parental figures lose some of their omnipotence. Piaget indicates that the change to logical-causal thinking becomes evident during the 7th or 8th year. More generally this period can be seen as a time of great upheaval. In our culture it is the beginning of formal education, the emergence from the family cocoon. The changing demands are likely to put the existing self-picture to severe test and as a result the concept of non-self or death is more apt to be responded to with increased emotional intensity.

---

[6] Significance values for t are estimated on the basis of a one-tailed hypothesis.

The interval from the ninth through the twelfth year has been labelled the pre-adolescent period, a time of latency. Measured in terms of psychological stress it is generally regarded as being a rather benign age. Rôles are well defined. Aggressive outlets are sanctioned by the culture in games and activities. Complex skills develop. In short no great new demands calling for marked change in response patterns are introduced. It is, as life goes, a "Golden Age." During this period questions about death disappear from the conscious repertoire as though this matter is no longer of interest. Children at this age seem to be too much involved in the routine of life and its attendant pleasures to be concerned with the concept of death.

The adolescent years in western culture have been written about as a period of "storm and stress." The individual is called upon to be an adult at one moment, a child in the next. There is little consistency in what the culture expects of the individual. Biologically he is an adult with no social status to match. This age (our 13 through 16 group) is characterized by the search for identity. The changing and often conflicting demands made on the adolescent provide the occasion and necessity for reorganization of the self concept. As in the 5 through 8 age period we may consider adolescence a nodal or crisis point in the life history, a time of a less stable self-picture and consequently a time when death has a greater emotional impact on the individual. This same argument, we hold, would apply to other "critical" points throughout life: marriage, menopause, old age, and states of pathology. In all these instances radical changes are demanded and death is likely to assume increased importance in the psychology of the individual.

What we are saying is that there is nothing peculiar to childhood per se that calls forth increased emotional response to death. It is rather that some periods of childhood are subjected to rapid and massive increases in psychological stress and as a consequence the decrease in ego stability is likely to be reflected in increased affective responses to death.

E. Summary

Affective responses to death words were measured in a population of males from 5 through 16. A word association task was employed in which response times and changes in skin resistance were used to indicate affect. Three age sub-groups are defined: 5 through 8, 9 through 12, and 13 through 16. The population as a whole shows increased emotional involvement with death words on both response measures. All sub-groups showed significantly increased response time to death words. On the *GSR* measure, two sub-groups, 5 through 8 and 13 through 16, show significant decreases in skin resistance. No reliable differences on this measure are found in the 9 through 12 group. The results are discussed in terms of cultural expectations and ego stability.

References

1. Alexander, I. E., Colley, R. S., & Adlerstein, A. M. Is death a matter of indifference? *J. of Psychol.*, 1957, **43**, 277-283.
2. Anthony, S. The Child's Discovery of Death. New York: Harcourt, Brace, 1940.
3. Bitterman, M. E., Krauskopf, J., & Holtzman, W. H. The galvanic skin response following artificial reduction of the basal resistance. *J. Comp. & Physiol. Psychol.*, 1954, **47**, 230-234.
4. Caprio, F. S. A study of some psychological reactions during pre-pubescence to the idea of death. *Psychiat. Quart.*, 1950, **24**, 495-505.
5. Chadwick, M. Notes upon the fear of death. *Int. J. Psychoanal.*, 1929, **10**, 321-334.
6. Cousinet, R. L'idee de la mort chez les enfants. (The idea of death in children.) *Psychol. Abst.*, 1940, **14**, 499 (*Abstract*).
7. Harnik, J. One component of the fear of death in early infancy. *Int. J. Psychoanal.*, 1930, **11**, 485-491.
8. Katz, D. Gespräche mit kindern. Berlin: Springer, 1928.
9. Kotsovsky, D. Die psychologie der todesfurcht. (The psychology of the fear of death.) *Psychol. Abstr.*, 1939, **13**, 134 (*Abstract*).

10. NAGY, M. The child's theories concerning death. *J. Genet. Psychol.*, 1948, **73**, 3-27.
11. OSIPOV, N. W. Strach ze smrti (Fear of death). *Psychol. Abstr.*, 1935, **9**, 534 (*Abstract*).
12. PIAGET, J. The Language and Thought of the Child. London: Routledge, 1952.
13. RASMUSSEN, V. Child Psychology (Vol. 11). New York: Knopf, 1922.
14. ROSENSZWEIG, S., & BRAY, D. Sibling deaths in the anamnesis of schizophrenic patients. *Arch. Neurol. & Psychiat.*, 1943, **49**, 71-92.
15. SCHILDER, P., & WECHSLER, D. The attitudes of children towards death. *J. Genet. Psychol.*, 1934, **45**, 406-451.
16. SULLIVAN, H. S. The Interpersonal Theory of Psychiatry. New York: Norton, 1953.
17. SULLY, J. Studies of Childhood. New York: Appleton, 1914.
18. WILK, M. B., & KEMPTHORNE, O. Fixed, mixed, and random models. *J. Amer. Statist. Ass.*, 1955, **50**, 1144-1166.
19. Teaching Spelling: Course of Study and Manual. Board of Education for the City of New York, Curriculum Bull., 1953-54 series, No. 6.

Department of Psychology
Princeton University
Princeton, New Jersey

The Journal of Genetic Psychology, 1958, **93**, 179-183.

CONDITIONED RESPONSES IN THE OPOSSUM*

Department of Psychology, University of Georgia

W. T. JAMES

A. INTRODUCTION

A previous study on the opossum indicated the animal could not develop a specific conditioned avoiding response of the foreleg to a bell and shock (1). Three hundred applications of the signal failed to form a conditioned avoiding response, although there were definite conditioned breathing changes and bodily movements to the stimulus.

An attempt was also made to develop a conditioned running response by sounding a signal and shocking the animal on the feet. Rather than run off the grill as one would expect, the animal gave a passive defense response (curling over on side and becoming inactive) which protected it from the grill.

During the past few years opossums have been reared in the Psychology Laboratory of the University of Georgia, and experiments on these animals indicate they can learn more than we expected. The animals learn to operate the Guthrie-Horton Puzzle box in a short time (2). They have also learned to operate various types of levers to open doors; to discriminate between cards on two doors; to solve the Fink-Arrow Maze, and to give operant responses in a lever pushing apparatus. All of these results have made it difficult to understand the earlier study. The animals used in the first experiment were wild. They would respond with the passive "defense reflex" when disturbed. They were active and would eat only during night time. The animals reared in the laboratory however never exhibit the passive defense reflex. They eat in the presence of the experimenter. When placed in the learning situations they go to work on the problem and are not distracted by the experimenter or laboratory environment. Since these animals are so much more responsive than the wild animals, the studies on conditioning have been repeated. It was thought possibly the wild animals were so disturbed during the tests, a true picture of conditioning in the opossum could not be obtained.

---

*Received in the Editorial Office on August 8, 1957.

B. Experiment 1

1. *Apparatus*

In the first experiment an attempt was made to form a conditioned avoiding response of the left rear leg to a shock. A restraining apparatus was constructed to confine the animal, yet allow freedom of movement of body and legs. The restraining frame was two feet long and 11¾ inches high, with a bar across the top 10 inches from the base. A yoke was made with a 2½ inch hole to fit around the animals neck. The yoke could be placed around the neck and locked into slots 7 inches from the front of the frame. The rear legs were supported by a harness suspended from the bar above (Figure 1). In this way the animal could move the body and legs, but it could not bite the electrodes or pneumograph. In transporting the animal from the home cage to the apparatus, the yoke was attached around the neck and the animal was carried by holding the yoke in one hand and the tail in the other.

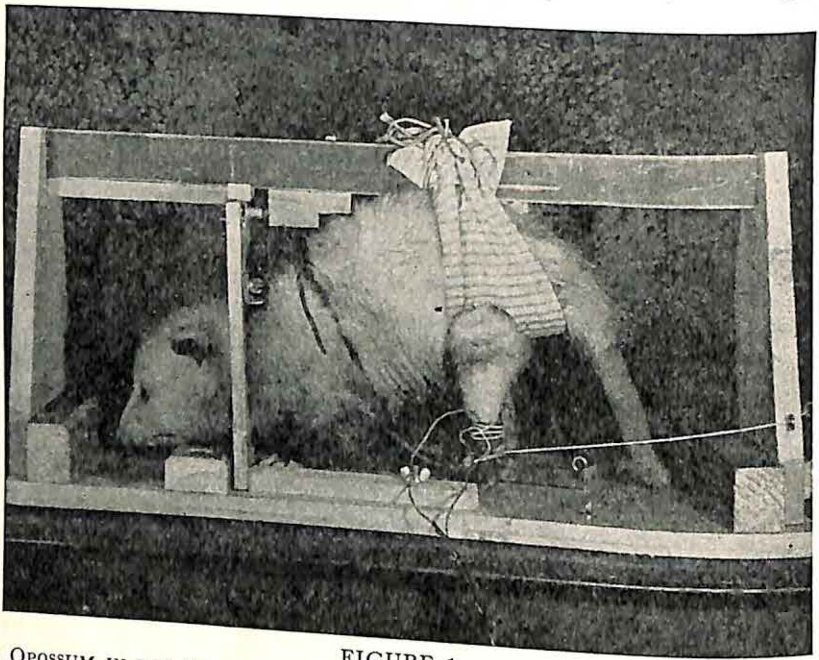

FIGURE 1

The Opossum in the Supporting Frame with Attached Pneumographs, Electrodes, and Recording String

2. *Procedure*

The procedure followed was to sound a buzzer from 1 to 5 seconds and present a brief shock on the rear left leg. Twenty-five signals were given each day, with an occasional signal of from 10 to 20 seconds, to determine if the animal would respond without shock. Kymographic records were taken periodically to determine the nature of breathing and leg response to the signal and shock.

3. *Results*

In the first study over 300 signals were applied. In the present experiment 550 signals were given without the development of a specific conditioned response of the leg. In the last 25 signals a total of seven conditioned reactions did occur, but as in the previous study, the leg action was always accompanied by total bodily action. Conditioned breathing also occurred to the signal, as in the previous case. A typical record, to Signal 545, is shown in Figure 2. The exaggerated breathing indicates the total bodily movement, superimposed on breathing. This experiment therefore seems to agree with the previous one. It seems to be impossible to form a conditioned response of one leg segment in the opossum. The action is always accompanied by total bodily movement.

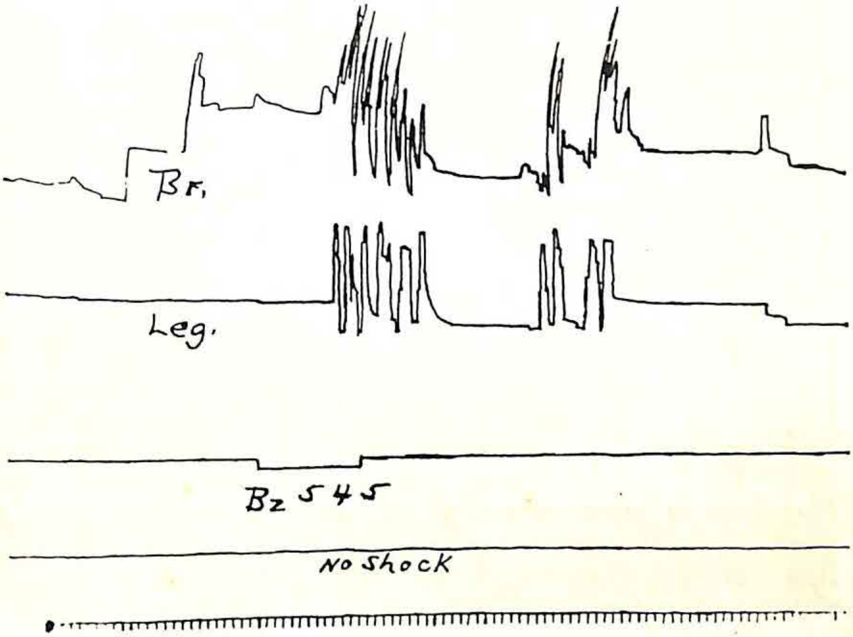

FIGURE 2
A Kymograph Recording to Signal 545: Breathing Changes Occur First, Followed by Flexion of the Left Rear Leg Superimposed on a "Walking" Pattern of All Legs

C. EXPERIMENT 2

1. *Apparatus*

Two boxes, each 20 by 8 by 12 inches, were placed end to end with a sliding door between them. A grill was placed on the floor of the left box. The animal could escape into the second box as the door was opened and shock applied to the bottom of the feet. Each side had a hinged top. A timing device was devised to start a clock as the conditioning signal was applied, and stopped as the animal jumped into the second box, giving an accurate measure of reaction time. The procedure followed was to sound the buzzer from 1 to 5 seconds then raise the restraining door and allow the animal to escape into the second compartment. Since the opossum is a rather slow moving animal, signals of long duration were given occasionally to test for the response.

2. *Results*

Three animals have been trained with this procedure. Animal 2 gave the first startle response to the buzzer on the seventh signal. The first series of 10 responses to the buzzer occurred after 53 signals. The average delay of these responses was 4.08 seconds. Animal 3 gave the first conditioned response after three signals, and 10 escapes in succession after 28 signals. The average delay was 4.98 seconds. Animal 4 gave the first conditioned response after five signals, and a series of 10 straight conditioned responses after 17 signals. The average delay was 2.72 seconds.

D. EXTINCTION

Since the opossum is a rather slow and inhibited animal it was of interest to observe the development of extinction. After the application of 63 positive signals in experiments with Animal 2, the shock was eliminated. This animal failed to respond to this first signal without shock, and extinction appeared immediately—10 signals of 10 to 20 seconds duration, with door opened between the boxes, without escape. Animal 3 was more active and signals of short duration were given without shock. After six signals of short duration the conditioned response had disappeared. The animal failed to respond to signals of long duration after this. Animal 4 gave eight responses to the first 20 signals without shock (short duration of signal), and then no response to the next 10 signals.

E. SUMMARY AND DISCUSSION

Since the opossum has failed to develop specific leg conditioning in two studies, it would seem we are justified in concluding this animal cannot form a conditioned response of one leg without having the excitation spread to all leg segments. The opossum is in a sense a spinal animal. The rear legs do not have a motor representation in the cortex (4). The forelegs have representation, but even in this case specific leg action could not be formed (as for example in the dog). The total escape pattern, however, is readily conditioned, although the reaction time is relatively long, in comparison to other animals.

The fact that this animal is inactive and inhibited fits in rather well with the readiness to form extinction. As soon as the shock is eliminated the response disappears. It seems to make sense that this inactive animal forms extinction within a short time. It has been found that extremely excitable dogs never form complete extinction of an avoiding response to a signal once it has been accompanied by shock (3). The more inhibited dogs do form extinction, and in general the more inhibited they are the quicker extinction forms. Since the opossum is in a sense further down on the continuum of inhibition, it has less difficulty in forming negative responses.

REFERENCES

1. JAMES, W. T. An experimental study of the defense mechanism in the opossum, with emphasis on natural behavior and its relation to mode of life. *J. Genet. Psychol.*, 1937, **51**, 95-100.
2. ———. The behavior of the opossum in the Guthrie-Horton puzzle box. *J. Genet. Psychol.*, 1955, **87**, 203-206.
3. ———. Morphological and constitutional factors in conditioning. *Ann. New York Acad. Sci.*, 1953, **56**, 171-183.
4. LANGWORTHY, O. R. Correlated physiological and morphological studies of the development of electrically responsive areas in the cerebral cortex of the opossum. *Contrib. Embryol.* (Carnegie Institute), 1927, **19**, 98-117.

Department of Psychology
University of Georgia
Athens, Georgia

The Journal of Genetic Psychology, 1958, **93**, 185-190.

KINESTHETIC CUES IN THE DEVELOPMENT OF TIME CONCEPTS*[1]

Baylor University College of Medicine and Houston State Psychiatric Institute

SANFORD GOLDSTONE, WILLIAM K. BOARDMAN, AND WILLIAM T. LHAMON

A. INTRODUCTION

The maturation of concepts involving duration, sequence, and order is of considerable interest to the student of development. Although attention[2] has been paid to the relationship between time concepts and mental age (2, 5) the emphasis has been upon the ability to use temporal language (6), understand the clock (10), and estimate durations and intervals of five seconds to several minutes under a variety of situational conditions (1, 3). Little work has been done on the development of specific subjective units of duration. Standard units of time permit calibration of personal and relative time in terms of a consensually valid frame of reference. Presumably one must reach a particular maturational level before culturally determined time standards may be learned. Recently Smythe and Goldstone (9) studied the passive estimation or judgment of tones as *more* or *less* than a clock second from which a subjective second was derived. It was found that children from age 8 through 14 did not differ from adults in their judgment of *seconds*. These children acquired the same average Second Estimation Point (*SEP*) as did adults and were able to correct their judgments with information as to the absolute value of the clock second. The only change from 8 to 14 years was a progressive decline in intra-subject variability of judgment suggesting the stabilization of the concept. On the other hand, the 6- and 7-year-old *S*s differed from the older children in their notions of a second, were extremely variable in their judgments, and were unable to profit from information. The average subjective second obtained with this passive method was consistently and significantly less than the standard clock unit. This suggested that the passive judgment did not permit the utilization of cues relevant to the learning and accurate appreciation and production of standard clock units. Since other studies (11) have demonstrated greater accuracy when the task involved counting seconds out loud,

---

*Received in the Editorial Office on August 20, 1957.

[1] This research was supported by Grant M-1121 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

[2] The literature on development and time perception was reviewed by Smythe, E. J. (8).

it was suspected that kinesthetic cues may be involved in the accurate learning and production of seconds. Preliminary observation suggested that the length of the subjective second was directly proportional to the amount of motor participation. The present investigation compared the accuracy of estimates of a second using the counting procedure with that of passive estimates in children from age 6 through 14, young adults, and older adults. The following questions were considered pertinent: (*a*) Is the Second Estimation Point (*SEP*) more accurate and longer when kinesthetic cues are available? (*b*) Does the counting method produce the abrupt maturation point at age 8 obtained with the passive method? (*c*) Does counting seconds aloud produce a longer *SEP* than counting to one's self with less muscle involvement?

B. Method

1. *Subjects*[3]

One hundred ninety white children of average or bright normal intelligence, as determined by group intelligence tests, were studied. Groups consisted of 20 *S*s (10 boys and 10 girls) at each age level from 6 through 13 and 30 *S*s age 14 (10 girls and 20 boys). In addition, 20 college students with a median age 24.0 and 20 healthy older adults with a median age of 69.5 years were studied. All *S*s except the college student group were tested with the passive method described elsewhere (9). Passive estimations of a young adult group were already available (4).

2. *Procedure*

S was instructed to count 30 secs. at the rate of one count per sec. under two conditions: (*a*) *Self*—*S* was asked to count 30 secs. to himself without tapping his fingers or feet and without vocalization. A starting signal was provided and *S* raised his hand to indicate completion. Five trials were obtained. (*b*) *Aloud*—*S* was asked to count 30 secs. out loud without restrictions upon tapping. Five trials were obtained.

Performance was timed with a stop watch to 0.1 sec. and the average time of the five trials was obtained. The average 30 secs. was divided by 30 to approximate an estimate of *S*'s concept of one second or counting *SEP*.

C. Results

Table 1 provides the median *SEP* values of the two counting conditions and their respective quartile deviations. The median *SEP* values of the passive estimation method are included to permit examination, but since these data are reported in detail elsewhere (9) no statistical elaboration will be provided here. Discrepancies in *N* reflect the number of *S*s that were able to complete each time estimation task. One *S* (6-year-old) was unable to complete both counting conditions whereas 9 six year *S*s, 6 seven year *S*s, 2 eight year *S*s, and 1 nine year *S* were unable to complete the passive method.

---

[3] We are indebted to Dr. H. M. Landrum, Superintendent of the Spring Branch, Texas, School District and the principals and teachers of the Spring Branch schools for providing *S*s and space for this investigation. Dr. E. J. Smythe collected the data.

TABLE 1

MEDIANS AND QUARTILE DEVIATIONS FOR THE COUNTING AND PASSIVE * *SEP* VALUES FOR ALL AGE GROUPS

	Counting to self			Counting aloud			Passive estimation*		
Age groups	N	Median *SEP*	Q	N	Median *SEP*	Q	N	Median *SEP*	Q
6	20	.61	.23	19	.59	.14	11	.84	.34
7	20	.63	.30	20	.61	.54	14	.41	.27
8	20	.99	.32	20	1.06	.24	18	.38	.23
9	20	1.00	.38	20	1.25	.22	19	.51	.34
10	20	.92	.35	20	1.09	.48	20	.48	.29
11	20	1.06	.21	20	1.11	.23	20	.46	.25
12	20	.96	.17	20	1.04	.22	20	.51	.16
13	20	.98	.11	20	1.08	.15	20	.50	.18
14	30	.93	.17	30	1.06	.24	30	.58	.13
Young adult	20	1.01	.18	20	1.06	.16	41	.56	.20
Older adult	20	.81	.19	20	.80	.17	20	.55	.35

*These results previously reported in Reference 9.

SEP differences among age groups for each counting condition were examined with the Mann-Whitney *U* test (7). Differences between the two counting methods for each age group were studied with the Wilcoxon sign-rank test (12). All significant results of these tests are summarized in Table 2.

2. *Counting to Self*

Tables 1 and 2 indicate that the *SEP* values obtained by the 6, 7, and older adult groups were smaller and less accurate than those provided by all other age levels. The median *SEP*'s obtained by counting to the self from 8 through young adult were remarkably accurate ranging from .92 secs. to 1.06 secs., whereas the passive median *SEP* reported elsewhere (9) and shown in Table 1 ranged from .38 secs. to .58 secs. *Q* values showed a slight decline from 11 through young adult indicating a more stabilized group norm or concept.

3. *Counting Aloud*

The *SEP* values obtained by the 6, 7, and older adult groups were again smaller and less accurate than those provided by all other age groups. The median *SEP*'s ranged from 1.06 secs. to 1.25 secs. for the 8 through young adult groups indicating considerable accuracy in contrast to passive estimation. Again the *Q* values were somewhat smaller from 11 through young adult.

TABLE 2
CONFIDENCE LEVELS

Age		8	9	10	11	12	13	14	Young adult	Combined 8-young adult
		Mann-Whitney U tests of significance among age groups for the two counting methods								
6	Self	X	.02	.10	.002	.05	.01	.01	.002	
	Aloud	.05	.002	.02	.002	.02	.002	.01	.002	
7	Self	X	.10	X	.01	X	.10	.10	.05	
	Aloud	X	.10	X	.10	X	X	.1	.10	
Older adult	Self	X	.05	X	.002	.10	.02	.05	.01	.002
	Aloud	.05	.01	.10	.01	.02	.01		.01	.002
6 & 7 combined	Self									.002
	Aloud									.002

Wilcoxon Sign-Rank test of significance between the two counting methods for each age group

Age	6	7	8	9	10	11	12	13	14	Young adult	Older adult
	X	X	.01	.01	.01	X	.01	.01	.01	.02	X

These data verify the abrupt change in the development of a standard time concept at about 8 reported elsewhere (9). The counting methods and their accompanying kinesthetic cues are apparently much more accurate than passive, non-muscular methods of estimation.

4. *Comparison between the Two Counting Methods*

A matched pair comparison of the *SEP*'s obtained counting aloud and to the self for each age group revealed significant differences at ages 8, 9, 10, 12, 13, 14, and the young adult group. Counting aloud produced a longer *SEP* for each age group and again this difference does not appear until age 8. The increased *SEP* was not simply a function of increased accuracy since as many *S*s became less accurate as became more accurate with this method.

It is of interest to note that many of the 6 and 7 year *S*s had considerable difficulty complying with the *counting to self* condition. They persisted in moving their lips and tapping their fingers, almost as though this kinesthetic activity was necessary for the expression of the temporal concept.

D. Discussion

This study demonstrated considerable accuracy of individuals from 8 through young adulthood in designating their concept of a standard clock unit—the second, by counting. Younger children and older adults were unable to display this accuracy in counting seconds. Presumably the basic mechanism in a counting procedure involves kinesthetic cues and we might assume that the learning of this short standard unit of time is at least in part muscular. Passive estimation of seconds provided through auditory stimuli when no motor production was involved resulted in a much shorter subjective second (4, 9). It is suggested that at about age 8, kinesthetic cues become associated with standard temporal concepts providing part of the mechanism for temporal conceptualization and calibrated experienced duration. This cut-off point at age 8 compares with previously reported results (9) and suggests this age as the approximate level of maturation which would permit temporal learning, and calibrated temporal experience. Further evidence of a kinesthetic base for learned temporal units was suggested by the younger children's difficulty in counting to themselves. It was necessary for many of them to violate instructions and make manifest muscle movements. It was as though these temporal concepts were in the process of being learned and needed the kinesthetic accompaniment. Also, most *S*s who took the passive test reported trying to employ a counting method to establish a standard for comparison. If this formulation is valid, we must presume that our 6- and 7-year-old *S*s had not as yet established a stable relationship between a standard duration and a muscle cue while the older adults apparently partially lost this relationship, thus accounting for their similarity to the younger children. Evidence from other studies (4, 9) suggests that auditory cues independent of motoric accompaniment provide little basis for accurate assessment of standard units of duration. It would be of interest to investigate the rôle of visual and tactile experience on temporal learning and appreciation.

The results of the comparison between counting to the self and aloud suggests that starting at age 8 the length of the subjective second is a function of the amount of muscle activity. Although both counting procedures involved muscle activity, more manifest action occurred when counting aloud, at which time a longer *SEP* was obtained.

If we regard learned units of our culturally determined dimensions as social norms which vary within limits with our frame of reference and permit calibrated, consensually valid interpersonal transactions, then time estimation measurements might be expected to vary with disorders of social norms. It would be of interest to compare the *SEP* values obtained here and with the passive method (4, 9) with those of children who demonstrate problems of social maturation manifested by behavior disorders, neurotic coping techniques, and psychoses. It would also be of interest to study the relationship between mental age and appreciation of standard duration units.

E. SUMMARY

One hundred and ninety children age 6 through 14, 20 young adults, and 20 older adults were asked to count 30 secs. aloud and to themselves at the rate of one count per sec. *Ss*' concept of a second was derived from these counts and comparisons were made among age levels and between the two counting methods. Considerable accuracy was obtained in the estimates of a second by the 8 through young adult groups. The estimates of 6 and 7 year and older adult groups were significantly shorter. Counting aloud, which involved more muscle activity resulted in significantly longer estimates of a second. These results are discussed in light of the rôle of kinesthetic cues in the learning and appreciation of standard units of duration. The relationship between active and passive estimations is also discussed.

REFERENCES

1. AXEL, R. Estimation of time. *Arch. of Psychol.*, 1924, **12**, 1-77.
2. BUCK, J. N. The time appreciation test. *J. Appl. Psychol.*, 1946, **30**, 388-398.
3. GILLILAND, A. R., & HUMPHREYS, D. W. Age, sex, method, and interval as variables in time estimation. *J. Genet. Psychol.*, 1943, **63**, 123-130.
4. GOLDSTONE, S., LHAMON, W. T., & BOARDMAN, W. K. The time sense: Anchor effects and apparent duration. *J. of Psychol.*, 1957, **44**, 145-153.
5. GOTHBERG, L. The mentally defective child's understanding of time. *Amer. J. Ment. Def.*, 1949, **53**, 441-453.
6. HARRISON, M. L. The nature and development of concepts of time among young children. *Elem. Sch. J.*, 1934, **34**, 507-514.
7. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122-143.
8. SMYTHE, E. J. A normative genetic study of the development of time perception. Unpublished doctoral dissertation, University of Houston, 1956.
9. SMYTHE, E. J., & GOLDSTONE, S. The time sense: A normative genetic study of the development of time perception. *Percept. & Mot. Skills*, 1957, **7**, 49-59.
10. SPRINGER, D. Development in young children of an understanding of time and the clock. *J. Genet. Psychol.*, 1952, **80**, 83-96.
11. STEINBERG, H. Changes in time perception induced by an anaesthetic drug. *Brit. J. Psychol.*, 1955, **46**, 273-279.
12. WILCOXON, F. Some rapid approximate statistical procedures. American Cyanamid Co., 1949.

Department of Psychiatry
Baylor University College of Medicine
Texas Medical Center
Houston, Texas

The Journal of Genetic Psychology, 1958, **93**, 191-198.

CONCEPTUAL BLOCK-SORTING AS A FUNCTION OF TYPE AND DEGREE OF MASTERY OF DISCRIMINATIVE VERBAL RESPONSES*

Department of Psychology, University of Massachusetts

ALBERT E. GOSS AND MARIE C. MOYLAN[1]

A. INTRODUCTION

Recent stimulus-response analyses (1, 3, 4, 6, 8) of the conceptual behavior of humans have included both task and verbal response-produced stimuli. In only a few studies, however, has there been direct, systematic observation or manipulation of mediating verbal processes. Analysis of overt verbalizations which occurred during presentation of Heidbreder's stimuli led Baum (1) to conclude that order of attainment of concepts varied inversely with number of names given to successive instances of the same concept. Fenn and Goss (4) observed that both normal and paranoid schizophrenic adults, who had learned nonsense syllable or familiar word labels which discriminated among blocks belonging to four height-size categories, subsequently sorted those blocks into height-size categories with fewer errors than controls. In a related study (2) with nursery school children, prior learning of familiar word labels for blocks resulted in more rapid learning of reinforced sorting by height-size. Acquisition of nonsense syllable labels, however, perhaps because they were less well retained or transferred, had no facilitative influence on learning to sort. Dietze (3) has reported that pre-school children who used dissimilar labels for figures of three different categories of shape of tip and shade of red acquired concepts more readily than another group of children whose verbal labeling responses produced highly similar stimuli.

In general, these results support a hypothesis that acquisition of discriminative labels for stimuli in different categories results in increased intra-category and decreased inter-category similarity of stimuli which, in turn, facilitates subsequent conceptual placement into those categories. Quantitative data describing the influence of potentially significant parameters, however, are lacking. Accordingly, the present study investigated transfer to sorting blocks into height-size categories as a function of degree of mastery

---

*Received in the Editorial Office on September 9, 1957.

[1] This study was financed in part by a grant-in-aid from The Society of Sigma Xi.

of discriminative familiar word or nonsense syllable labels. Both types of labels were used in order to ascertain whether this factor also influenced subsequent sorting behavior.

Acquisition of discriminative labels has been interpreted (4) as establishing an "abstract set" (5) to sort according to particular categories. A group instructed to sort by height-size was included, therefore, to compare the relative effectiveness of different degrees of strength of the presumed abstract set resulting from experimental acquisition of labels with that induced by experimental instructions to sort by height-size.

B. METHOD

1. *Subjects*

Within the restriction of half males and half females in each group, 150 undergraduates in an introductory psychology course were assigned to 15 groups of 10 *S*s each.

2. *Stimuli and Apparatus*

The 16 blocks represented combinations of tall (1 in.) or short (½ in.), large (0.80 sq. in.) or small (0.43 sq. in.), and square or circular top and bottom areas, and black or white. *Tall-large*, *tall-small*, *short-large*, and *short-small* were the familiar word labels for blocks in those categories and *bik*, *cev*, *lag*, and *mur* were the corresponding nonsense syllable names.

The blocks were presented one at a time in the window of an exposure box, which, along with the sorting board, have been described elsewhere (4).

3. *Procedure for Learning Labels*

Familiar word or nonsense syllable labels for blocks were learned by the paired-associate technique. Each block was presented alone for a two-second interval at the end of which a bulb on the top of the box lighted and *E* told *S* the familiar word or nonsense syllable name of the block. After *S* had repeated the name, the window was closed and another block was substituted. All 16 block-label pairs were presented in different random orders within successive cycles of 16 trials each.

Successive degrees or criteria of mastery of both types of labels are shown in the left-hand column of Table 1. These were one correct anticipation for any single block within any one of the four height-size categories (one), two correct anticipations for any two blocks not in the same category (two), three correct anticipations for any three three blocks not in the same category (three), one block in each of the four categories anticipated correctly (four), and correct anticipations for nine (*9/16*) and 14 (*14/16*) of the 16 blocks in four successive sequences of four blocks each.

TABLE 1
MEANS AND STANDARD DEVIATIONS OF NUMBER OF VERBAL LEARNING TRIALS AND CORRECT ANTICIPATIONS DURING THOSE TRIALS

	Type of label							
	Familiar words				Nonsense syllables			
Degree of Mastery	Trials		Correct anticipations		Trials		Correct anticipations	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
One	10.3	2.5	1.0	0.0	15.6	7.5	1.0	0.0
Two	15.7	4.2	2.9	0.7	22.7	13.0	2.3	0.6
Three	13.4	5.3	3.4	0.6	28.3	10.6	4.0	1.2
Four	27.5	13.1	6.8	2.4	28.1	6.8	6.9	1.4
9/16	29.3	12.4	12.1	5.8	60.3	22.7	15.3	5.1
14/16	39.0	21.6	20.5	10.7	78.8	26.0	34.7	8.3

In addition a *see-discriminate-name* (*SDN*) condition was introduced to control for exposure and attending to the blocks as well as naming, verbal categorizing, and "learning set" experiences with them. All *S*s in this condition were given 76 trials which approximated the 78.8 required for the 14/16 nonsense syllable group to reach criterion. During these 76 trials each block was exposed for approximately the same length of time as blocks were seen in each verbal learning trial. The remaining two of the 15 groups were the *instructed* condition and *controls* neither of which saw the blocks before block-sorting.

4. *Procedure for Block-Sorting*

Each of the four block-sorting trials began with placement of four blocks—one for each of the height-size categories—in outside corners of the four squares of the sorting board, one block to each square. The remaining 12 blocks were then exposed one-by-one for *S*s to place each one in the corner which contained a matching block.

Different sets of four blocks, one for each category, were assigned to the four squares on the four successive block-sorting trials. There was no limit on time taken for placement, but correction of placements was not permitted nor were *S*s told whether or not placements were correct. Inter-block intervals were about five seconds and inter-trial intervals about 30 seconds.

The 12 verbal learning groups and the *SDN* group were given block-sorting instructions immediately after completion of presorting experiences. In addition to these instructions *S*s of the *instructed* group were told to sort the blocks by height-size and, more specifically, into *tall-large*, *tall-small*, *short-large*, or *short-small* categories.

C. Results

1. *Learning Labels*

With the exception of two inversions, number of verbal learning trials and correct anticipations during those trials increased for successively higher correct anticipation criteria (Table 1). Familiar words were learned more rapidly than nonsense syllables. Also, the relative advantage of the former appeared to increase as criteria became more stringent.

2. *Block-Sorting*

Means and standard deviations of numbers of 48 possible correct height-size placements for all four block-sorting trials combined are shown in Table 2 and Figure 1. *Control* and *SDN* groups sorted the blocks by height-size at chance levels. All of the six *familiar word* groups made more correct placements than *control* and *SDN* conditions. There was, however, at most only a slight upward trend from the one correct response criterion to the 14/16 criterion with inversions at the four correct and 9/16 points. All of the *nonsense syllable* groups also made more correct placements than *controls* or Ss of the *SDN* group. However, learning nonsense syllables to one, two, or three correct criteria produced little if any advantage in block-sorting. The upward trend for the higher levels of mastery of nonsense syllable responses was marked but, even at the 14/16 point, fell below the number of height-size placements of the group instructed to sort by height-size.

TABLE 2

Means and Standard Deviations of Correct Placements for Combined Block-Sorting Trials

Degree of mastery	Type of label Familiar words Mean	SD	Nonsense syllables Mean	SD
One	22.7	14.7	13.9	14.0
Two	22.4	14.7	14.9	13.7
Three	25.4	14.0	13.0	12.9
Four	18.7	19.5	22.9	15.4
9/16	22.8	20.1	23.8	14.6
14/16	27.9	17.6	36.9	13.7
Control	7.1	6.3		
Instructed	43.9	7.8		
See-discriminate-name	12.8	13.4		

Twelve points for the verbal learning groups, all *above* the *control* group mean and all 12 points *above* the mean for the *SDN* group, would occur by chance with a frequency of 1/4096. Twelve points all *below* the mean of the *instructed* group would be equally rare. Therefore, statistical analysis indicates that learning familiar words or nonsense syllables produced more height-size placements than seeing-discriminating-naming or no exposure to the blocks. Specific appropriate instructions, however, resulted in even more height-size placements.

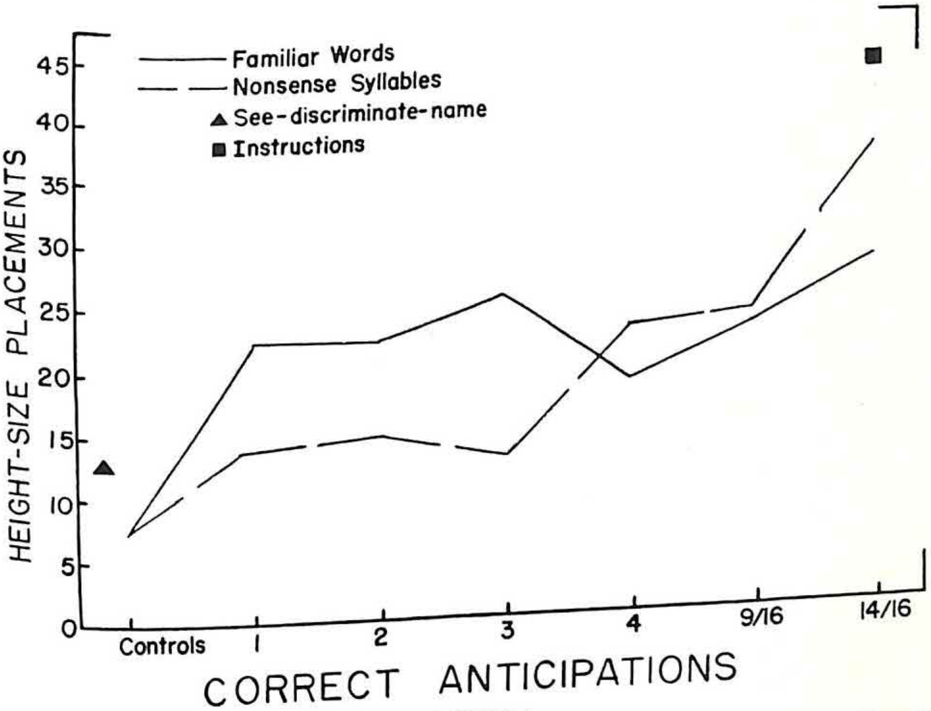

FIGURE 1
NUMBER OF CORRECT HEIGHT-SIZE PLACEMENTS FOR COMBINED BLOCK-SORTING TRIALS AS A FUNCTION OF DEGREE OF MASTERY OF VERBAL RESPONSES

Table 3 summarizes a Lindquist (7) mixed design (type III) analysis of means of correct placements for each of the four block-sorting trials (not shown to conserve space) for seven criteria of mastery of familiar word or nonsense syllable labels. For this analysis, *S*s of the *control* group were divided into two equal subgroups which had approximately equal mean correct placements on each trial. One subgroup was then assigned to control or zero level of mastery of familiar words and the other to control or zero level of mastery of nonsense syllables. Only main effects of degree of mastery and of test trials were significant. The former indicates that number of correct placements increased with greater mastery of labels. The latter was due to somewhat less frequent correct placements on the first block-sorting trial. Although the trends for familiar words and nonsense syllables appeared to be different the interaction of degree of mastery and type of label was not significant.

TABLE 3
ANALYSIS OF VARIANCE OF CORRECT RESPONSES FOR BLOCK-SORTING TRIALS

Source	*df*	Mean square	*F*
(T) Type of label	1	42.12	——
(D) Degree of mastery	6	206.61	3.25*
T x D	6	73.66	1.16
Error (between)	116	63.50	——
(TT) Test trials	3	34.13	16.98*
T x TT	3	1.88	——
D x TT	18	2.45	1.22
T x D x TT	18	2.58	1.28
Error (within)	348	2.01	——

*Significant at < .01 level.

D. DISCUSSION

Acquisition of familiar words or nonsense syllables which differentiated height-size categories increased subsequent sorting in terms of those categories. The one correct anticipation criterion may have produced some facilitation which did not increase markedly out to the three or four correct anticipation points. From those points on, at least for nonsense syllables, there was evidence of an upward trend to the 14/16 criterion. The apparent initial plateau suggests that for lower degrees of mastery sorting may have been by height *or* size, but not by height *and* size. The curve of correct sorting responses for familiar words did not exhibit a pronounced upward trend beyond the three or four correct anticipations criteria. The nonsignificant interaction of type of label and degree of mastery, however, suggested that this apparent difference in trends may have been due to chance factors.

With a somewhat different, more stringent 14/16 verbal learning criterion Fenn and Goss (4) obtained slightly more correct placements than occurred for familiar words or nonsense syllables learned to the 14/16 criterion of this experiment. It seems probable, therefore, that the asymptote for transfer to block-sorting lies somewhat beyond the 14/16 point of Figure 1.

More precise specification of the form of the relationship between height-size placement and degree of mastery is difficult because the degree of mastery verbal learning criterion employed is not unequivocally unidimensional. However, degree of mastery was related by upward trends to the "single" dimensions of number of verbal learning trials and of correct anticipations.

Instructions to sort by height-size led to even more correct placements than acquisition of labels to the 14/16 criterion. Because a block from each of the relevant categories was on the board, it is assumed that instructed *S*s inspected the blocks on the board and made correct pre-experimentally acquired labeling responses to each. Then, as blocks were presented in the window, the instructions assured that the same labels occurred, and that the exposed blocks were placed with those on the boards which had the same names. Greater mastery of verbal labels is apparently required to produce a sorting performance "set" of effectiveness equal to that induced by instructions. Or such effectiveness may be contingent on *S*s being told to use experimentally strengthened verbal responses to aid their sorting.

E. SUMMARY

Number of placements of blocks by height-size was investigated as a function of degree of mastery of familiar words or nonsense syllables which discriminated among the four height-size attributes of 16 blocks representing combinations of tall or short, large or small and square or circular top and bottom areas, and black or white. Including 0-level of mastery controls, there were seven degrees of mastery. One of two additional groups looked at, discriminated among, and named the stimuli without differential reinforcement; the other had no presorting experience with the blocks before being specifically instructed to sort by height-size.

Extent of facilitation apparently increased as a function of degree of mastery, with nonsense syllables producing a lower initial but higher subsequent trend than familiar words. Instructions, however, resulted in the largest number of correct placements. Interpretation was in terms of learning-induced or instruction-induced verbal mediating responses whose stimuli presumably increased intra-category and decreased inter-category similarity.

REFERENCES

1. BAUM, M. H. A study in concept attainment and verbal learning. Ph.D. dissertation, Yale Univ., 1951.
2. CAREY, J., & GOSS, A. E. The rôle of mediating verbal responses in the conceptual sorting behavior of children. *J. Genet. Psychol.*, 1957, **90**, 69-74.
3. DIETZE, D. The facilitating effect of words on discrimination and generalization. *J. Exp. Psychol.*, 1955, **50**, 255-260.
4. FENN, J. D., & GOSS, A. E. The rôle of mediating verbal responses in the conceptual sorting behavior of normals and paranoid schizophrenics. *J. Genet. Psychol.*, 1957, **90**, 59-68.

5. Goldstein, K., & Scheerer, M. Abstract and concrete behavior; an experimental study with special tests. *Psychol. Monog.*, 1941, No. 239.
6. Kendler, H. H., & Vineberg, R. The acquisition of compound concepts as a function of previous learning. *J. Exp. Psychol.*, 1954, **48**, 252-259.
7. Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. New York: Houghton-Mifflin, 1953.
8. Osgood, C. E. Method and Theory in Experimental Psychology. New York: Oxford Univ. Press, 1953.

Department of Psychology
University of Massachusetts
Amherst, Massachusetts

The Journal of Genetic Psychology, 1958, **93**, 199-228.

RAVEN'S PROGRESSIVE MATRICES: A REVIEW AND CRITICAL EVALUATION*

VA Hospital, East Orange, New Jersey

HENRY R. BURKE[1]

The Progressive Matrices test, described by its author, J. C. Raven, as a test of "innate eductive ability" (87), is not widely known or used in the United States. Two major versions of the test have been published: Progressive Matrices (1938) (83, 84) and Progressive Matrices (1947) (90, 96).

A. PROGRESSIVE MATRICES DESCRIBED

1. *Progressive Matrices (1938)*

Progressive Matrices consists of 60 perceptually presented tests. Each test consists of a design or "matrix" from which part has been removed. The testee has to examine the matrix and decide which of the pieces given below is the right one to complete the matrix. Twelve tests complete a Set and there are five Sets, lettered *A* to *E*. Each Set develops a different theme (87, p. 12).

The first problem in a set is intended to be self-evident, and it is succeeded by 12 problems of increasing difficulty. . . . The themes employed were (*a*) continuous patterns, (*b*) analogies between pairs of figures, (*c*) progressive alterations of patterns, (*d*) permutations of figures, and (*e*) resolution of figures into constituent parts.

For children from 3 to 8 years of age and for the mentally defective, Sets *A* and *B* were arranged as coloured patterns on cardboard (Board form of the test). The act of filling up the gap in the board trained the child effectively. For normal subjects over the age of 8 years, the matrices in Sets *A*, *B*, *C*, *D*, and *E* were drawn in black and white and mounted (Standard form of the test). The number of problems correctly solved represented the subject's total score. . . . For self-administration or group testing, the instructions were modified: the alternatives were numbered from left to right in each problem and scoring forms were provided. Subjects were told how to solve the first problem (Set *A*, No. 1) and how to enter their choice on the scoring form: afterwards they worked steadily through the series at their own speeds and without further assistance (Group form of the test). The problems are arranged so that success on the easy ones at the beginning helps to train the subject to solve the more difficult ones which come later: verbal instructions were, thus, reduced to a minimum, and a fixed form of repetition was not insisted upon. The psychologist made as little comment as possible and recorded the results (86, pp. 16-17).

---

*Received in the Editorial Office on September 11, 1957.

[1] The reviewer wishes to express his thanks for the advice and encouragement of Doctor Laurance F. Shaffer of Teachers College, Columbia University, and Doctor Henry Kavkewitz of the VA Hospital, East Orange, New Jersey.

2. *Progressive Matrices (1947), Sets A, Ab, B*

In 1947, Progressive Matrices (1947), Sets *A, Ab, B* (Coloured Progressive Matrices) (96), was constructed: to give, for children of 5 to 11 years of age, a wider dispersion of scores, to reduce the frequency of chance solutions and to make the test more suitable for use with persons who are for any reason mentally subnormal or impaired. For this purpose, a transitional set of 12 problems is placed between Sets *A* and *B* of the 1938 scale. The problems in this set are constructed to be intermediary in difficulty between Problems 5 and 12 of Set *A* and Problems 1 and 7 of Set *B*. They are arranged so that, for the three sets combined, children between the ages of 5 and 11 solve about three additional problems each year.

To attract and hold the attention of little children, each problem is printed on a brightly-coloured background. This makes the nature of the problem to be solved more obvious without in any way contributing to its solution. The order of the problems in each set provides the standard training in the method of working and the three sets together are arranged to cover all the perceptual reasoning processes of which children under 12 years of age are usually capable. If the test is suitably presented, it is necessary only to show a person what to do, to let him work through the problems in the standard order and to learn from his own experience to solve them. The test can be presented in the form of boards and movable pieces, or as illustrations printed in a book, without the intellectual processes required for success being essentially altered. In either form, the problems to be solved can be demonstrated quite satisfactorily without any verbal instructions at all (97, p. 1).

3. *Progressive Matrices (1947), Sets I and II*

A revised and more difficult form of Progressive Matrices grew out of its use with the British Services in World War II (36). Progressive Matrices (1947), Sets I and II (90) was constructed: for use with persons over 11 years of age of average or more than average intellectual ability. They can be used without a time limit in order to assess a person's maximum capacity for observation and clear thinking, or with a time limit to assess his speed of accurate intellectual work. In the first set, there are 12 problems. They are designed to introduce a person to the method of working. In the second set, there are 48 problems. In presentation and argument, they resemble the problems in Sets *C*, *D*, and *E* of the 1938 scale. So that the validity of the total score does not depend upon everyone attempting all the problems in this scale before stopping, they are arranged in groups of 4 according to argument and order of difficulty. An adult of superior intellectual capacity takes at least half-an-hour to solve them all, while a person's interest and attention are usually maintained for more than an hour without boredom. Set I will indicate in a few minutes whether a person can be regarded as intellectually "dull," "average," or "bright." If a person appears to be of average or more than average ability, it can be followed by Set II and the estimate made more exact (85, p. 3).

B. History and Use of *PM* (1938)

1. *Origin and Use in World War II*

Progressive Matrices, previous to its development by Penrose and Raven (81), and Raven (82), had its origins (7, 60) in the thinking and work of Burt, Spearman, and their students, e.g., Stephenson (120), Line (66). The use of *PM* (1938) for batteries with British Army recruits in 1940 was said (7) to have been suggested by Burt. It has been, perhaps, the most widely used of British tests (20, 110, 125, 133), e.g., taken by some three million recruits in the last war (126). It is now seldom given in the Services, mainly because its wide use has caused it to become rather well known (133). It has had easier testee acceptance than verbal tests in clinics and the Services because of its apparent nonverbal character and lack of educational bias (133).

2. *Review of the Literature*

In its almost 20 years of existence there has been no critical review of the literature. This review has listed more than 50 English, 10 American, and 10 non-English substantial studies of *PM* (1938). A search of the literature has found only three English studies (33, 34, 94) of *PM* (1947), Sets I and II, and four American studies of Coloured *PM* (1947) (40, 68, 118, 119).

3. *Studies With Special Groups*

Progressive Matrices (1938) has been used in published studies of many special groups: the deaf (63, 76, 77, 78, 111), the cerebral palsied (50, 121, 123), mental defectives (12, 16, 47, 55, 56, 57, 75, 90, 122), adult psychiatric patients (13, 15, 18, 20, 21, 25, 27, 28, 29, 30, 31, 32, 42, 43, 44, 48, 49, 61, 108), child guidance clinic patients (3, 56, 57, 87), army recruits (1, 5, 19, 37, 104, 105, 109, 124, 125, 126, 127, 128, 129, 33, 138), university or professional students (10, 22, 33, 36, 39, 45, 46, 51, 52, 62, 64, 65, 67, 71, 79, 80, 112, 139, 140), apprentices (9, 80).

C. VALIDITY

What is actually measured by Progressive Matrices (1938), its validity, is not easily settled.

1. *Opinion*

a. Raven's "eductive ability." Opinion begins with statements by the author: "*By itself* it is not a test of 'general intelligence' " (85). He has spoken of it as a "means of estimating a person's innate eductive ability" (87), as designed to measure "accuracy of education" (87), as covering "the whole range of eductive ability" (87), as "a test of a person's present capacity to form comparisons, reason by analogy, and develop a logical method of thinking regardless of previously acquired information" (95), as "designed . . . the best test for this purpose yet produced" (91). Raven (24) offered a definition of intellectual ability as the "ability to reason by analogy from awareness of relations between experienced characters."

b. Spearman's g. Spearman (113, 114, 115) considered Raven's Progressive Matrices a test for measuring *g* or eduction, in fact, perhaps the best of all nonverbal tests of *g* (116, 117). This has been a general English opinion, e.g., Vernon (127) spoke of it as "one of the purest tests of *g* available," Vincent (134), "an almost pure *g* test." The evidence on this point from factor analyses will be considered later.

D. CONTENT VALIDITY

1. *Evidence from Correlation Data*

a. PM (1938) and the Binet test. Raven (95) reported a correlation of .86 between *PM* scores and Terman mental ages for 150 normal school children, with an age range of 6 to 13. Taibl (121), with 63 selected spastic subjects for whom *T-M IQ*'s were available in the approximate age range of 7 to 15, reported a correlation of .72 with *PM* (1938) Grades

I-V. Keir (60), in a careful and critical study, cited from a study by Horton (1939) a correlation of .62 between scores on *PM* and Burt's revision of the Terman-Binet, for 123 boys and girls, ages 11 to 13. Banks and Sinha (7), also, in another critical study of *PM* (1938) gave a partial correlation of .54 between *PM* and the Terman-Merrill Revision of the Binet, with age partialled out, for children ages 8 to 11, (N ?). Johnson (55), in a not very well designed or controlled study with a training school population of 30 boys and 30 girls, with a mean age of about 12, reported barely significant correlations of .45 and .41 between *PM* (1938) and the Binet.

b. PM (1938) and the Wechsler test. In a recent study in the United States, Barratt (8), with 70 children, ages 9.2 to 10.1, found the following correlations betwen *PM* (1938) and the *WISC*: Total Score .75, Verbal .69, Perf. .70. Coefficients of correlation between Wechsler-Bellevue Verbal and *PM* (1938) scores are available for psychiatric patients: mean age 27 (*N* 94), .47, Halstead (44); mean age 61, (*N* 35), correlation with age partialled out, .38, Bromley (13); mean age 31.7, range 16 to 63 (*N* 190), .57, Desai (21); age range 19 to 49 (*N* 120), .64, Klonoff (61). Klonoff (61) also found a correlation between *PM* (1938) and *W-B* Total Scores, $r = .74$; for *W-B Perf.* Scores and *PM* (1938) $r = .71$. For college students, mean age 19 (*N* 60), Levine and Iscoe (62) gave a correlation of .55 for *W-B* Total Score (Short Form) and *PM* (1938). The same authors (62) reported a questionable correlation of .80 between *PM* (1938) scores and mutes, age ranges 15 to 19 (*N* 73).

c. PM (1938) and performance or nonverbal tests. Bolton (11) reported a questionable correlation of .80 between *PM* (1938) scores and Arthur Perf. *MA*'s for 23 non-English speaking East Chicago elementary school children, ages unspecified. On the basis of this correlation and criteria of time and practicality, in 1950, he used *PM* (1938) in place of the Arthur with this type of population.

The same author (11) for the Pintner Non-Language test, with children in Grades 4 to 12, ages unspecified, reported correlations with *PM* (1938): .53 (*N* 290), and .51 (*N* 204). Levine and Iscoe (63) with 73 deaf children, ages 15 to 19, reported a correlation of .41 between Chicago Non-Verbal scores and *PM* (1938). Barratt (8) found a correlation of .58 for 70 fourth grade children, with an age range of 9.2 to 10.1, between *PM* (1938) and Columbia Mental Maturity scores. Rimoldi (100), with 138 randomly selected school children, ages 11 to 14, found the following correlations between scores on Porteus Mazes and *PM* Sets: *A*, .36; *B*, .35;

C, .36; *D*, .48; *E*, .22. Tizard *et al.* (122), with 140 high grade mental defectives, mean age 20.9, *SD* 4.6, reported a correlation of .54 between Porteus Mazes *IQ*'s and "estimated" *PM* (1938) *IQ*'s. Emmett (23) with 178 primary school boys, ages 11-0 to 12-11, found correlations of .51 and .73 between two nonverbal mental ability tests and *PM* (1938). Moore and Peel (71), with 40 volunteer third- and fourth-year dental students found a correlation of .61 between *PM* (1938) and Peel's Adult Non-Verbal test.

d. PM (1938) and verbal group tests. Correlations of *PM* (1938) with other group tests of intelligence have been reported: Bolin (10), for 76 junior college students, mean age 18.7, with Otis Gamma, .65; Bolton (11), for 83 school children, ages unspecified, with Otis Gamma, .40; Germain (38) for a homogeneous group of 27 students, with Otis, .67. Bolin (10) also reported correlations of the *ACE* with *PM* (1938): Total Score .48, *L* Score .29, *Q* Score .59. Bolton (11), for school children, ages again unspecified, reported correlations of *PM* (1938) with the Henmon-Nelson (*N* 154), .47, and the Terman-McNemar (*N* 164), .58. For 600 male adult soldiers in service, Rudolf (105) found a correlation of .27 between Kent *EGY* Test Scores and *PM* (1938). With the Cattell Scale III, largely verbal, on a volunteer population of 95 medical students, Gibson (39) found a correlation of .30. Moore and Peel (71) working with 40 volunteer third- and fourth-year dental students, reported a correlation of .48 between scores on the Cattell Scale III and *PM* (1938). Tizard *et al.* (122) reported a correlation of .61 between *PM* (1938) "estimated *IQ*'s" and Cattell scores for 104 adult high grade mental defectives, mean age 20.8, *SD* 4.6.

Halstead (42, 44), for adult psychiatric patients (*N* 1326 and 94), reported correlations of .55 and .47 between *PM* (1938) and Shipley Vocabulary scores. For 75 senile patients, Eysenck (29) reported a correlation of .26 between *PM* (1938) and Mill Hill Vocabulary scores. Between *PM* (1938) and Mill Hill Vocabulary scores, Raven (95) found the following correlations: .57 (age 13 plus or minus 1, *N* 41), .60 (age less than 21, *N* 48), .51 (ages 30 to 39, *N* 37), .45, (ages 40 to 48, *N* 35), .44 (age 50, *N* 29). Keir (69), for her population of school children, ages 10 to 13, reported a correlation of .41 for 128 boys and a correlation of .36 for 168 girls between *PM* (1938) and *MHV*.

e. PM (1938) and memory tests. Eysenck and Halstead (28), in a factorial study of 15 memory tests with 60 male Army neurotic patients, found the correlation of these tests with *PM* (1938) to range from .96 to

.63. Banks (5), in a factorial study with British Army recruits (578 *M*, 595 *F*), found correlations of .64 and .54 between a memory test and *PM* (1938). These data, incidentally, support the opinion of Bradford (12) that "a good memory for patterns" or "the memory factor" plays a very important rôle in this test. Rimoldi (100) also considered Set *E* "loaded in a memory factor."

f. PM (1938) and spatial tests. Emmett (23), for two tests of spatial relations in two and in three dimensions, for which he claimed an odd-even corrected reliability coefficient of .97, found correlations (with age partialled out) of .62 and .41 with *PM* (1938) (*N* 178 school boys, ages 11 to 12-11). Bernyer (9), in a study of 160 select male apprentices in a school for mechanics, age range 13.5 to 17, reported a correlation of .38 with scores on *PM* (1938) and the Minnesota Paper Form Board. Moore and Peel (71), with 40 volunteer dental students found a correlation of .27 with the National Institute for Industrial Research Form Relations test. Halstead (44), with 94 adult psychiatric patients, reported a correlation of .30 between *PM* (1938) and the Minnesota Spatial Relations test. It may well be that what *PM* (1938) measures in the perception of spatial relations differs at different age levels.

2. *Summary*

No studies report correlations of *PM* (1938) with other tests of mental ability as high as that given by Raven (95), .86 with the Terman-Merrill. The combination of verbal and performance scores on the Wechsler tests gave the next highest correlations with *PM* (1938), e.g., .75 (8) and .74 (61). Correlations between scores on *PM* (1938) and performance tests, and perhaps on nonverbal tests of mental ability and memory tests, are higher than correlations of *PM* (1938) and verbal group mental ability tests. Correlations of *PM* (1938) with vocabulary test scores are not very different from, but perhaps lower than, correlations with verbal group mental ability test scores. The inter-test correlations are of course functions of the age range and homogeneity of the subjects as well as of the test reliabilities, which, in the case of *PM* (1938), especially for the part scores, and in narrow age ranges, may be lower than is commonly assumed by those who use *PM* (1938).

E. CONCURRENT AND PREDICTIVE VALIDITY

1. *Correlation With Educational Achievement*

Correlations of *PM* (1938) with achievement, not unexpectedly, vary at least with the way achievement is measured, with sex, with the reliability of *PM* (1938) at different score levels, and possibly with the age at which achievement is measured.

Controlled-answer testing: (*a*) with Army recruits in Banks' (5) study, gave correlations of .69 (men), .49 (women), with Arithmetic; .63 (men), .41 (women) with Bennett's Mechanical Comprehension test; (*b*) with 178 11-year-old school boys in Emmett's (23) study, gave correlations of .51 with English, .48 with Arithmetic problems, .34 with Arithmetic computation; (*c*) in Keir's (60) study with 10- to 13-year-old school children, with Arithmetic .36 (128 boys), and .31 (168 girls).

Estimates of achievement or course grades at the university level showed the following correlations with *PM* (1938): Looze (67) reported biserial correlations on a pass or fail criterion and *PM* (1938) for university students: .56 (*N* 242), .57 (*N* 366).

Banks (5), with British Army recruits, found correlations of .56 (578 men) and .36 (595 women) between *PM* (1938) and educational level. Adkins and Lyerly (1), in a factor analysis study with 200 American Army recruits in 1950, ages 18 to 23 and educational range 6 to 14 years of school, found correlations with education and *PM* (1938), Set *B*, .22; *C*, .35; *D*, .26; *E*, .36. Bolton (11), for 888 school children, Grades 4 to 12, ages unspecified, reported a correlation of *PM* (1938) with age, .68.

2. *Prediction*

Harding (45) found *PM* (1938) "surprisingly useful . . . extremely promising in predicting success for students of architecture."

Edholm and Gibson (22) for "volunteer" medical students showed correlations between *PM* (1938) and physiology grades: .23 (*N* 90), .38 (*N* 81), .33 (*N* 77). The use of the criteria employed in this work has been criticized by Heim and Timpany (46), who also noted an error in the statistics. Orton and Martin (79) have used *PM* (1938) in screening medical students. Moore and Peel (71) found *PM* of some value in predicting aptitude for dentistry. Houliston (51, 52) reported that *PM* (1938), in conjunction with the Mill Hill Vocabulary Scale, was very useful in eliminating unsuccessful nursing applicants.

Bolton (11), with elementary school children's averages for a year in the fourth grade and *PM* (1938), found the test of little predictive value, correlations of .39 (*N* 219), .41 (*N* 283), .14 (*N* 212).

Vernon (126), discussing research on personnel selection in the British services, reported what he called "validity coefficients" for *PM* (1938), correlations between test scores and proficiency. In all cases these correlations were higher for individuals at the 90th percentile on *PM* (1938), .65, .62, .45; than (*a*) for individuals at the 50th, .49, .40, .28; or (*b*) at the 10th, .27, .26, .10. These data are related to the evidence of Roberts (104), as reported by Vernon (125), on the lower reliability of *PM* (1938) in the lower score ranges. Vernon and Parry (133) also indicated these coefficients were lower for *PM* (1938) than for an arithmetic, clerical, and vocabulary test. Vernon (130), in a popular article, asserted that a test battery of a nonverbal test, e.g., Raven's *PM*, one or two verbal and arithmetic tests, a mechanical, and a spatial judgment test "gives a useful prediction capacity for almost any service job."

3. *Summary*

The correlation data thus far summarized indicate fairly good, but not outstandingly superior, concurrent and predictive validity for *PM* (1938). More evidence may be marshalled from the available data.

F. CONCURRENT VALIDITY

1. *Standardization Data*

Confirmatory evidence for the validity of *PM* (1938), in the sense of capacity to differentiate among age groups below 15, has been furnished in graphs and tables provided in Raven's (86, 88) standardization data on Colchester and Ipswich children, age 13 to 14, also in the work of Rimoldi (100, 102, 103) with Argentine and American subjects, age range 7 to 15. Scatter diagrams for Raven's *PM* Grades I-V and Terman-Merrill results (*N* 301) (*N* 131), without coefficients of reliability, were also provided by Raven (85, 87). Walton (136) has published a similar table, from which he concluded that *PM* (1938) is not a nonverbal equivalent of the Terman-Merrill.

Evidence for the validity of *PM* (1938), in the sense of capacity to discriminate among groups known by other criteria to differ in intellectual capacity, is available. Raven's (86) early study with Colchester children showed discrimination between mean scores of five school groups which differed on experienced teachers' estimates of ability. In the same author's (88) Ipswich study, Graphs I and II showed the distribution over the entire range of scores for 1,407 children and 3,665 adults. *PM* (1938) discriminated among the mean scores: (*a*) of children at different educational levels, and (*b*) of children of parents in different occupational groups.

Delys (19) very neatly presented tables of norms as evidence for the validity of *PM* (1938) in discriminating between Belgian military recruits:

(*a*) from different language and socio-economic backgrounds, and (*b*) from different educational levels. (Norms for British Army recruits on *PM* (1938) have evidently not been published, although Heim and Wallace (47) mentioned norms on a representative Army sample of 5,000 made available to them by Ungerson.)

2. *Data from Atypical Groups*

Oleron (76, 77, 78) has used *PM* (1938) to demonstrate the marked inferiority of the deaf as compared with the hearing.

Spastic cerebral palsied have been shown by Taibl (121) to differ significantly in *PM* (1938) scores from the athetoid cerebral palsied and normal children. Tracht (123), on the basis of his experience, recommended the use of *PM* (1938) with the cerebral palsied. Holden (50), however, said its use with spastics needed further evaluation.

The capacity of *PM* (1938) to distinguish mentally defective children and adolescents is demonstrated in the study of Raven and Waite (99) and Tizard *et al.* (122). Davidson (18) presented evidence for differences in mean scores by normal, manic depressive, and schizophrenic groups.

Notcutt (73) has used *PM* (1938) to demonstrate a retardation of over two years among South African Zulu primary school children in comparison with English norms. He has published another study (74) with this group which was unavailable for inspection.

3. *Occupational Group Data*

Raven (88), in his Ipswich study, found *PM* (1938) to discriminate among the mean scores of adults in different occupational groups. This finding has been confirmed in other work of Raven (95), in Foulds and Raven (34, 35), Foulds (33) and Vernon (127, 128). Vernon (128) has published a table of occupational norms for the 20-minute *PM* (1938). Research in the British Armed Services, as reported by Vernon (126) and Vernon and Parry (133), has also confirmed this finding. Vernon, however, made the observation that "individual differences within occupational groups are much larger than differences between these groups" (127, p. 58). He stated that on the basis of Matrix scores, 78 per cent of the variance was due to individual differences, 13 per cent to occupational differences.

4. *PM (1938) Scores and Age*

The evidence accumulated by Bradford (12), M. Eysenck (29), Foulds (33), Foulds and Raven (35), Raven (95), Slater (109), Vernon (126, 127), Vernon and Parry (133), and Vincent (134) indicates that scores on *PM* (1938) reached a maximum at about 14 to 17, held constant until about 25, and that the regression of age on scores was then almost linear.

a. Correlation evidence. This has been contradictory. Keir (60), in her careful study, reported a correlation of .11 between *PM* (1938) and age for 296 children between the ages of 10 and 13. Eysenck (29, 30) with 75 senile patients, mean age 73.4, *SD* 6.5, found a correlation of age with *PM* (1938), Sets *A* and *B* only, of .04, below even the 5 per cent level of significance. Bromley (13), with 35 "primitive" psychiatric patients, mean age 61, reported a coefficient of .46 between age and *PM* (1938). Vernon (126) reported a correlation of .24 between *PM* (1938) scores and age for 90,000 young naval recruits from a dozen broad occupational groups. Vernon and Parry (133), for 578 Army recruits, ages 18 to 40, reported a correlation with increasing age from 30 to 50, .93 to .83, and declining intercorrelations 95) have also indicated a declining test-retest reliability for *PM* (1938) with increasing age from 30 to 50, .93 to .83, and declining intercorrelations of Mill Hill Vocabulary Scale with *PM* (1938), .60 to .44, (*N* 331).

5. *Growth and Decline of Intelligence*

Raven summarized his findings as follows:

> The capacity to form comparisons and reason by analogy increases rapidly during childhood, appears to have reached its maximum somewhere about the age of 14, stays relatively constant for about 10 years, and then begins to decline, slowly but with remarkable uniformity. This finding agrees with the findings of other investigators. . . .
>
> If the decline continues at the same rate after the age of 60, it would appear that by the age of 80, the average person's capacity to succeed in the Matrices test is less than that of the normal child of 8 years . . . (95, pp. 15-17).

It is interesting to note that M. Eysenck (29) found the level of 75 seniles on *PM* (1938), mean age 73.4, *SD* 6.5, equal to that of 8-year-old children. Crown (17), in a preliminary study, has used *PM* (1938) and the Mill Hill Vocabulary Scale to measure intellectual deterioration and improvement after therapy with four myxodematous patients.

There seems to be some disagreement as to whether the evidence has indicated that the rapidity of decline of effective intelligence was associated with occupation and education, as Vernon(127), or was independent of the work engaged in, as Foulds (33), and Foulds and Raven (35).

a. Percentile ranks for IQ's. Raven (85, 89, 93, 95), on the basis of the evidence for the changing ratio between age and test scores on *PM* (1938), especially before age 14, has been opposed to converting *PM* test scores into *IQ*'s. He has preferred the use of percentile ranks. Jones (58) has argued for the conversion of *PM* (1938) scores into *IQ*'s. Raven has been supported by Slater (110) and Walton (137).

b. Summary. There is abundant evidence of concurrent validity for *PM* (1938), in the sense of its capacity to discriminate over a wide range among groups known by other criteria to differ in intellectual capacity.

G. FACTOR ANALYSIS AND CONSTRUCT VALIDITY

1. *PM and Spearman's g*

Spearman (115) himself seems to have set English opinion in regarding *PM* (1938) as a nonverbal test of *g*, "the best of all" (116, 117); to quote Vernon and Parry (133) as late as 1949, "an almost pure *g* test."

Vernon (125, 126, 129, 132) reported that several factor analyses of *PM* (1938) in the British Services in World War II found that the Matrix test had *g* and *k*: *m* loadings of about .79 and .15, accounting for about 62 and 2 per cent of the variance. Adock (2), reworking some of Slater's data, with *PM* (1938), found evidence of group factors *k* and *v* accounting for a very small per cent of the variance, in addition to a general factor, *g*, in Matrices. Emmett (23), with 178 boys at age 11 plus, found a general factor with a *PM* (1938) loading of .83, a verbal factor with a *PM* loading of —.01, and a *k* factor with a *PM* loading of —.07. Klonoff (61), in a factor analysis of the Wechsler-Bellevue and *PM* (1938) with a group of neurotic veteran patients, age range 19 to 49, *M* 28.2, *SD* 7, found an *R* factor (Relationship Eduction) in the Raven, similar to an *R* factor (Logical Reasoning) in the *W-B*. In his opinion, the evidence pointed to *PM* (1938) as a measure of pure *g*.

2. *Multiple Factors in PM*

Jastak (54), however, has expressed the suspicion that *PM* (1938) measures several important factors in addition to the general factor, and that to find the Matrix test a relatively pure test of *g* must be due to faulty experimental design.

Three factor analysis studies of *PM* (1938) alone have been made. In what seems to have been a very careful study of *PM* (1938) with 296 primary and secondary school children in the East End of London, ages 10 to 13-11 (128 *M*, 168 *F*), Keir (60) found three factors, the first of which accounted for 37 per cent of the total variance, the second for 7 per cent, and the third for 6 per cent. In the factor analysis in Sinha's (107) study of *PM* (1938), also reported by Banks and Sinha (7), with 310 children, ages 8 to 13 (140 *M*, 170 *F*), *g* accounted for 38 per cent of the variance. Vernon (131), in a study of *PM* (1938) with 640 recruits (ages close to 18), which he believed confirmed Keir's (60) study, reported a *g* factor accounting for 18 per cent of the variance, and group factors for 9 per cent. He felt that the figure for *g* might be more than doubled in a representative group.

3. *PM Not a Pure Test of g*

Some other factor analysis studies have also not been clear in identifying *PM* (1938) as a pure test of *g*. In a factor analysis of *PM* (1938) by Gabriel (37) with 830 male and female recruits in the Israel Army, the first component accounted for 60 per cent of the total variance, the second 15 per cent, the third 10 per cent, the fourth 8 per cent, the fifth 6 per cent. Gabriel wrote: "Progressive Matrices do not form a perfect Simplex . . ." (37, p. 12). Banks (4, 5), in a factor analysis of British Army recruits (*M* 578, *F* 595), done under Burt, found a general factor accounting for 41 per cent of the total variance, with *PM* (1938) loadings of .84 for the men, and .74 for the women, a second factor accounting for 7 per cent of the variance, with *PM* factor loadings of .07 for the men and —.34 for the women. Bernyer (9), in a study using 20 tests with 160 boys selected out of 482 by a scholarship examination, ages 13.5 to 17, found a *g* factor, with 10 tests having saturations greater than .40, also a first and second factor; *PM* loadings were: .35, .53, .30. He considered the *g* factor similar to the "mental energy" or *g* factor of Spearman, the first and the second factor to be similar to the Induction and Spatial factors of Thurstone. Rimoldi (101), in a factor analytic study with 138 Argentine children, ages 8 to 15, found an *A* factor with Raven subtest loadings of *D*, .68; *C*, .55; *B*, .54; *E*, .45; and *A*, .34, similar to Thurstone's *I*.

Adkins and Lyerly (1) did a very thorough and comprehensive factor analysis study of 66 variables (45 to 48 were *PM* (1938) subtests *BCDE*) with 200 American Army recruits, ages 18 to 23 and 6 to 14 years of schooling, 26-minute time limit for *PM* (1938) subtests. They found a *B′* factor (Hypothesis Verification), "some aspect of which is often referred to as inductive ability," with *PM* subtest loadings of *C*, .57; *D*, .54; *B*, .47; and *E*, 32; a *D′* factor (Perceptual Speed) with a *PM* subtest *E* loading of .32, and an *M′* factor (Concept Formation) with a *PM* subtest *E* loading of .28.

The authors reported no evidence for a "general reasoning" factor, nor "any factor fully corroborating Spearman's *g*" among their first order factors.

Tizard *et al.* (122), in a factor analysis study with 104 adult high grade mental defectives, mean age 20.9, *SD* 4.6, extracted only one factor among 18 variables, which they thought could not be identified with Spearman's *g*, but was instead a complex nonverbal practical factor unrelated to verbal intelligence.

4. *PM and a Memory Factor*

Eysenck and Halstead (28), in a factor analysis of 15 memory tests and *PM* (1938) with 60 selected neurotic British Army veterans, got one general factor accounting for 74 per cent of the variance, with a .94 factor saturation with *PM*, and offered proof that the factor being measured by the intelligence test was identical with the factor being measured by the 15 memory tests.

The evidence from M. Eysenck's (29) factor analysis study with 75 male senile patients is difficult to relate to other findings on the nature of the factors in *PM* (1938). In her opinion the evidence offered a picture of mental organization in seniles differing much from that found in normal adults (31).

5. *PM and Eduction*

Guilford (41), in attempting to formulate a system into which the intellectual factors already identified by factor analysis could be fitted, made 7 out of 40, memory factors: figural, structural, conceptual; and the remaining, thinking factors: cognitive, productive, and evaluative. He considered Raven's Progressive Matrices an example of tests of a figural cognitive (discovery) factor of "eduction of perceptual relations," different from a figural productive (convergent thinking) factor of "education of correlates."

It is also difficult to align the factor analysis data on *PM* (1938) unequivocally in support of the conclusion of Burt: "Biological, neurological, and statistical evidence all appears strongly to support the view that with increasing age intellectual ability tends to become more and more specialized" (14, p. 85).

6. *Summary*

The evidence is not convincing that *PM* (1938) has validity as a pure measure of the Spearman construct of *g*; and doubt may be raised whether such a construct can be measured independently of the modality through which it is expressed, the selectivity of the subjects and their sex, and possibly the presuppositions of the factor analyst (131).

H. ITEM ANALYSES

Keating (59), in 1949, in his review of *PM* (1938), was obviously unaware of the item analysis that Raven (86) mentioned he did in the preparation and revision of his test for publication. Miller and Raven (69) also have published a report of their study of the influence on choices of the position of the alternative answers in *PM* (1938). Halstead (42), in 1943, published a table giving the difficulty order of items in *PM* (1938) with 2,790 unselected adult subjects, showing that many of the items were out of the order of difficulty as determined by this group. M. Eysenck (30) found an order of difficulty similar to Halstead's ($p = .95$), with 104 children, ages 7 to 9, and 100 senile patients. Rimoldi (103) also claims a similarity to this order of Halstead in his item analysis of *PM* (1938) with 1,680 Argentine children, ages 9 to 15.

Keir (60), in her study with 296 school children, ages 10 to 13-11 (*M* 128, *F* 168), found a correlation of .98 for the order of difficulty of boys and girls, and a correlation of .97 with the order of difficulty she got from Slater (110), who found a "preponderance of easy items." Keir (60) noted her agreement with Raven's order was "not so close." She stated that there were too many items of medium difficulty, and that the items were far less evenly spaced than the items of the Binet scale. Keir concluded: "A study of the validity of the separate items is urgently needed. . . . With the present group the validity coefficients, assessed by biserial correlation, range from 0.28 to 0.62" (60, p. 147).

Banks and Sinha (7) and Sinha (107), in reporting a study of 310 children, ages 8 to 13 (*M* 140, *F* 170), with *PM* (1938), also noted a bunching of items in the middle of the scale. They reported biserial coefficients of correlation for pass or fail on each item and the average score on *PM* and Simplex, or on *PM* and *T-M*, ranging from .2 to .8, with an average of .45. They listed definite items with poor *r*'s at almost every age. They reported "in the main the items in Set *A* possess little discriminative value" (7, p. 92). Adkins and Lyerly (1) have noted the not unexpected nondiscriminative character of Set *A* with their 18- to 23-year-old Army recruit population. Banks and Sinha (7) also mentioned the "large number of comparatively worthless items in Sections *B* and *C*" (7, p. 92). As to the difficulty order of the items with their population, they wrote: "Many of the items are obviously out of place; others are logically defective" . . . (7, p. 93). The actual order of difficulty correlated only .48 with an order of difficulty suggested by a method proposed by Burt. There is a recent German reference

(106) to the order of difficulty of items in Set *E* with superior subjects; it was unavailable for inspection.

Summary. The available item analysis data indicate the necessity of improving the validity and reliability of *PM* (1938). Although not inadequate, careful cross-validation studies seem to indicate validity and reliability are lower with children than with adults, and in general lower than claimed.

I. RELIABILITY

Further evidence in regard to the reliability of *PM* (1938) remains to be evaluated. Desai (20) has published a neat summary of the reliability data on *PM* (1938) up to 1952. Raven (95) himself, in 1948, was of the opinion that more accurate information concerning the reliability of the test at different age and score levels was still needed. There is some evidence for the greater reliability of *PM* (1938) when given untimed (131) and individually (136), but these matters need to be more systematically investigated. Raven (97) has himself also noted the low reliability of results in the lowest score ranges. This low reliability was studied and reported by Roberts (104), and his conclusions, while not generally available, have been widely quoted (7, 125, 133).

1. *Reliability in the Services*

Vernon, on the basis of his experience with *PM* (1938) in the British Services (124), remarked about the "rather poor reliability" of *PM* (1938) (125). He (126) gave a number of test-retest reliability coefficients: .79, Navy men (*N* 500), 6 to 9 months interval; .88, Army men (*N* 500), 6 to 8 weeks interval; .87, *ATS* women (*N* 1,000), 4 weeks interval. Vernon and Parry (133) also gave a test-retest reliability coefficient of .79, for 537 seamen at a 1 to 6 months interval. Delys (19) reported on the use of *PM* (1938) with Belgian Army recruits, with both French and Flemish speaking men, that many calculations of coefficients of reliability always gave results greater than .85.

2. *Raven's Reliability Data*

Raven (85), himself, in his *Guide to Using Progressive Matrices (1938)*, gave no reliability coefficients, but only Table VI: "Relationship between Percentile Grade and Terman Merrill Intelligence Quotients for a clinic group of 301 children given each test individually" (85, p. 13). This table was evidently based on his standardization data published elsewhere (88). In his study with the *RECI* Series of Perceptual Tests, predecessor of *PM*

(1938), Raven (86) reported a number of test-retest reliability coefficients of correlation: for 56 normal children, 5 to 9 years old, retested on the Board Form, Sets *A* and *B*, .86; for 120 normal children, retested at a 3-week interval on the group test, Sets *A* to *E*, mean age 13.5, .88; for 308 normals, ages 3 to 9, with Sets *A* and *B* in Board Form, the correlation was "low." Raven (95) reported the following test-retest coefficients of reliability: .88, 13 plus (N 41); .93, under 30 (N 48); .88, 30 to 39 (N 37); .87, 40 to 49 (N 35); .83, 50 plus (N 29). Foulds and Raven (35) report for volunteer employees the following test-retest coefficients of reliability: under 30 (N 44); .93; 30 to 40 (N 104); .88; 40 to 50 (N 106), .87; 50 plus (N 77), .83.

3. *Reliability With Psychiatric Patients*

In his careful study with 300 male psychiatric patients, ages 18 to 65, mostly neurotic, Desai (20) obtained a test-retest coefficient of reliability after four weeks, .74. For *PM* Groups I to V he found different reliabilities: I (N 52), .57; II (N 68), .33; III (N 112), .27; IV (N 50), .41; V (N 18), .73.

With 100 neurotic patients H. Eysenck (26) obtained a test-retest reliability coefficient of .81, with *PM* (1938) untimed, after a six weeks interval; with a control group of Civil Defense Workers, the coefficient was .87. The same author (26), in a study with a population of 300, concluded that the intelligence test scores of neurotics were barely significantly less reliable than those of normals. Raven (89) cited test-retest reliability data from an evidently unpublished study by Halstead, for 107 neurotics, $r = .76$, for normals, $r = .88$.

M. Eysenck (29, 30) with 84 male senile patients, mean age 73, on the *PM* (1938) Board Form, Sets *A* and *B*, after a 4-month interval, got a test-retest coefficient of reliability of .49. She (32) has also found a test-retest coefficient of .76 between grades on two testings, after a two-year interval, for 39 epileptics, ages 5 to 19.

4. *Reliability Data in Critical Studies With Children*

Coefficients of reliability with 296 East End of London children, ages 10 to 13-11 (*M* 128, *F* 168), in Keir's study (60), were lower than those given by Raven (86). She (60) cited a reliability coefficient from an unfinished and unpublished study by Horton, for 123 boys and girls, ages 11 to 13.5, $r = .71$. Keir's (60) own data, using the split-half technique for odd and even items, duly corrected, gave .76. For test-retest reliability after a two-year interval with 41 11-year-olds, she reported a coefficient of .64.

Vernon (131), in a study with 640 recruits, in which he thought the results confirmed Keir's, reported the internal consistency of the test, shown by the *K-R* Formula 20, to be .85, compared with Keir's odd-even reliability coefficient of .76 with children.

Sinha (107) and Banks and Sinha (7), with 140 boys and 170 girls, ages 8 to 13, reported a split-half corrected reliability coefficient of .86, or a reliability coefficient of .88 for the average of: the .86, an analysis of variance coefficient of .90, and a Kuder-Richardson coefficient of .91.

Bolton (11), in a rather naïvely designed and reported study, without specifying ages or numbers, reported a split-half, corrected, coefficient of reliability for *PM* (1938), .90.

5. *Adequate Reliability Generally Assumed*

It has been interesting to see certainly not unsophisticated investigators like Adkins and Lyerly (1), Bernyer (9), Bradford (12), Emmett (23), Klonoff (61), Levine and Iscoe (63), Notcutt (73), Rimoldi (100, 101, 102, 103), Slater (109), Taibl (121) employing *PM* (1938) without apparent serious discussion of the reliability of Progressive Matrices and a check on the reliability of the test with the populations in their studies.

6. *Need for Improved Reliability*

In view of the data on the need for greater reliability, and the data on content, predictive, and construct validity of *PM* (1938), it is hard to agree with the statement of its author: "No general revision of it appears to be either necessary or desirable" (85).

Keir's (60) conclusion seems more in line with the evidence: "The figures obtained raise serious questions in regard to the reliability and efficiency of the test as a whole, and the value and arrangement of several items. These questions require to be satisfactorily solved before it can be accepted as a serviceable school test" (60, p. 140).

Sinha (107) concluded similarly: "Although the test was decidedly promising in its general lines, it would be premature to accept it in its present form, when there was so much room for obvious improvement" (107, p. 239).

7. *Summary*

The evidence is not convincing that the test *PM* (1938) has the reliability, at the various age levels, that is necessary and desirable for important decisions about an individual. Test-retest reliability coefficients with normal adults in their late teens or early twenties range from .93 (35) to .79 (26).

Pertinent reliability coefficients in careful studies with children range from .88 (86) to .71 (60).

J. Research Findings

1. *Use With Psychiatric Patients*

Research findings with *PM* (1938) not already indicated may be listed here. The test is reputed to be one of the most widely used tests with psychiatric populations in England (20). Davidson (19), with small groups, found progressive increases in difficulty with the test for normals, manic depressives, and schizophrenics. Eysenck (25), in a large study with some 3,000 neurotic patients, found evidence that neurotics did not on the average differ from normal people in their intellectual level, and support for the view that it was people of greater or less than average intelligence who became maladjusted to their environment and developed neurosis. In another study (26), he also found evidence to support the conclusion of Spearman that improvement on cognitive tests is relatively independent of conation. Slater (108), with *PM* (1938), found obsessionals more intelligent than other neurotics.

In a study of 168 outpatient epileptics with *PM* (1938), "a fairly representative group," M. Eysenck (32) found 12 per cent mentally defective.

Halstead (42) reported, in a study of 700 neurotics, that mere fluctuation of scores over the subtests of *PM* (1938) was not a criterion of neurosis. He also reported evidence from the same study that "speed" and "power" on mental tests were correlated. Using *PM* (1938), Himmelweit (48) found differences between patients with anxious and depressed makeup and patients with hysterical makeup. The same author (49), with similar groups, found hysterics showed a speed preference with *PM* (1938), and the dysthymics an accuracy preference.

Halstead and Slater (44) found *PM* (1938) to have poor predictive power for success in occupational therapy with neurotic patients. Johnson (56, 57) has some unconvincing reports on the use of Raven's *PM* (1938) with children to appraise "ego-capacity" or potential for progress in play therapy.

Smyth (112) summarized the findings of Leybourne-White (64, 65) with *PM* (1938) on the abilities of University of Manchester medical students in comparison with the abilities of the total university and general population: 6 per cent of the total population had ability equivalent to that of the average university student; 9 per cent of the population had ability equal to that of the average medical student. Only one in five in the general population with ability equal to that of the upper half of university students got to the university.

Warburton (139), for 862 college students, reported a non-significant correlation of .03 between *PM* (1938) scores and size of family, found by Leybourne-White (64). Warburton (140), for 481 technical college students, using *PM* (1947), found an effectively zero correlation with the same criterion.

2. *International Use of PM*

Raven (85) has called attention to the international use of *PM* (1938). Alon-Bakaliar (3) found it not to be fitted to the mentality of oriental communities in Israel. Wurburton (138), also, found it not to be useful with an isolated primitive group of Gurkas. These facts are perhaps evidence that the test is not entirely culture-free.

There is a French manual by Raven (98) cited in the literature, and a French standardization (*N* 1237) published in 1950 (72), which we have not been able to see. There are French studies using *PM* (1938) by Bernyer (9), Delys (19), Looze (67), Oleron (76, 77, 78). Studies of *PM* (1938) have also been published in Argentina, Rimoldi (102, 103); in Spain, Germain (38); in Italy, Casini-Nencini (15), Misiti (70); in Australia, Oxlade (80); in Germany, Seeger (106); in Israel, Alon-Bakaliar (3).

3. *PM Culture Free*

The evidence from these is not entirely clear that *PM* (1938) can be described as a culture-free test. Westby (142) has expressed his doubts. Bradford (12) has also expressed doubt about the independence of the test of schooling. Perhaps it is as nearly culture-free as any other available test is or can be. Conclusive evidence on the point is lacking.

4. *An American Matrices*

For Americans, attention might be called to a matrices test of American construction, the Navy-Northwestern-Matrices (53); comparative studies might be initiated.

5. *PM Reviewed*

Attention may also be called to the fact that two reviewers of *PM* (1938), Wechsler (141) and Westby (143) have written of the test's "main limitation," or "critical defect," as its attempt to measure with "a single modality of performance," or "one set of counters." Attention might also be called to the relatively small number of items.

K. Coloured Progressive Matrices (1947)

Wall (135), reviewing Progressive Matrices (1947), Sets *A*, *Ab*, *B* (96), Coloured Progressive Matrices, ages 5 to 11, untimed, 30 minutes, considered the test "still experimental," and "not a substitute for more thorough measurement." Banks (6) called attention to the fact that no item analysis had been published.

1. *For Use With Children*

Raven (97), in *Guide to Using Progressive Matrices (1947), Sets A, Ab, B, Revised*, described the test as "constructed to give, for children of 5 to 11 years of age, a wider dispersion of scores, to reduce the frequency of chance solutions and to make the test more suitable for use with persons who are for any reason mentally sub-normal or impaired." He claimed a test-retest reliability of .65 for children under 7; by age 9, a test-retest reliability of at least .80, and over the whole range for which the test was constructed a test-retest reliability of .9; for children less than 7, a correlation of .5 with the Terman-Merrill Scale, Form *L*. He provided tentative percentile norms based on Scottish Dumfries children, ages 5 to 11.5 (*N* 608).

2. *Use With Mentally Defective Children*

Stacey and Carleton (118) have used *CPM* (1947) with 150 children in a State School, with an age range of 7-5 to 15-9, Stanford-Binet *IQ*'s 39 to 89, *WISC* Full Scale *IQ*'s 36 to 91. They reported the following correlations with *CPM* (1947): *CA*, .54; *S-B* (*L*) *MA*, .69, *IQ*, .71; *WISC* Full Scale Score, .62, *IQ* .55; *WISC* Perf. *IQ*, .52; *WISC V IQ*, .54. Correlations with *WISC* subtest weighted scores ranged from .48 to .28.

3. *Use With Mentally Defective Adults*

Stacey and Gill (119) have also used *CPM* (1947) with 172 subnormal adults, 135 not institutionalized but on convalescent care, *CA* 16 to 57-6, *M* 23-10, *S-B IQ*'s 40 to 84; *M* 63.8, *W-B* Full Scale *IQ*'s, 46 to 102, *M* 69.7. The following correlations with *CPM* (1947) were obtained: *S-B* (*L*) *IQ*, .86; *W-B* Full Scale *IQ*, .68, *W-B* Perf. *IQ*, .51; *W-B V IQ* .56.

4. *Use With Normal Children*

Martin and Wiechers (68) used *CPM* (1947) with 100 Indiana school children between their ninth and tenth birthdays, *WISC* Full Scale *IQ*'s 74 to 141, Mean 107, *SD* 16.1. The following correlations were found with *CPM* (1947): *WISC* Full Scale *IQ*'s, .91, *WISC* V *IQ*'s, .84; *WISC*

Perf. *IQ*'s, .83. Correlations with *WISC* subtest scores ranged from .74 (Block Design) to .47 (Picture Arrangement).

a. Group testing. Green and Ewert (40) used *CPM* (1947) with 1,214 normal Minnesota children, ages 6 to 12.5, by means of group testing with colored slides, time limit 20 minutes. The Raven norms appeared too low. Correlations of *CPM* (1947) were obtained as follows: with Otis, for children ages 6-6 to 12-5 (*N* 192), .78; with California Mental Maturity Test *IQ*'s and Kuhlman-Anderson *IQ*'s, for small groups (*N* 22 to 51), at different age levels, $r = .28$ to .56. The procedure seems hardly to have been an adequate or useful trial of *CPM* (1947). The authors concluded that the test should not be thought of as a test of nonverbal reasoning ability but as a test of fairly complex intellectual reasoning processes.

b. Evaluation. *CPM* (1947) cannot be considered a substitute for the Stanford-Binet or the Wechsler tests but the correlations of these tests with *CPM* (1947) for these populations seem as high as the correlations reported for these tests with one another in similar populations. *CPM* (1947), in view of its ease of administration and the time required, seems worthy of further study, item analyses, cross validation studies, and normative data for special groups.

L. PROGRESSIVE MATRICES (1947), SETS I AND II

The use of Progressive Matrices (1947), Sets I and II (for adults, 40 minutes) (90), has not been reported beyond the two studies of Foulds (33) and Foulds and Raven (36), and an unpublished report by Raven (94). Foulds (33), comparing scores for university students on *PM* (1938) (*N* 880) and *PM* (1947) (*N* 170), showed that the "roofing" effect of *PM* (1938) with this group disappeared with *PM* (1947). The latter test gave an effective range four *SD*'s below and above the mean. Further concurrent validity data were presented in a table which showed the differences in mean scores and scatter for three occupational groups of different educational levels.

Test-retest reliability data were given: (*a*) for adult students (*N* 243), $r = .91$; for children, age 12.5 (*N* 92), $r = .86$; for children, age 10.5 (*N* 109), $r = .76$. Foulds considered the reliability of the test unsatisfactory below the 11-year-old age level.

> It appears to provide a valid and reliable test of intellectual efficiency . . . at the age of 11 years, and will differentiate clearly between adults of superior intellectual capacity (33, p. 245).

The Matrices test provides a reliable index of a person's present capacity for intellectual activity independently of the language he speaks or of the knowledge he has acquired. From a person's percentile grade and score composition for the Matrices test in relation to his score composition on the Vocabulary test, it is possible to assess the normality of his intellectual development, compare people of any education, and, for clinical purposes, infer the degree of intellectual deterioration which has occurred as the result of age or brain injury (33, p. 246).

Foulds and Raven (36), on the basis of item analysis data which they got from a study of 304 adults at a teacher training college, made recommendations for the revision of *PM* (1947). They presented reliability data already published by Foulds (33). The joint authors presented concurrent validity data in greater detail through the medium of score distributions of 1,344 subjects in groups differing in age and educational level, from 10-year-old suburban school children in a dockland area to university science majors, e.g., *M* 10.3, *SD* 5.1 (*N* 129), *M* 34.1, *SD* 3.2 (*N* 27). The authors presented comparative norms in terms of percentiles for *PM* (1938) and *PM* (1947): with 417 children, age 10.5 years; with 596 children, age 12.5 years; and with 304 20-year-old adults.

Summary. Cross validation studies are required to establish the reliability and validity of *PM* (1947), Sets I and II. We have data supplied only by the author and his colleague. It may well be, as they have asserted, that perceptual tests of education can serve as valid and reliable means for assessing intellectual capacity at the highest levels (94, p. 15).

M. MANUALS

A final criticism needs to be made of the manuals or guides for Progressive Matrices (85, 97). By American Psychological Association standards, summarized in *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (144), these manuals are deficient and inadequate. The extent of this inadequacy is apparent in a comparison of them with Wechsler's *Manual for the Wechsler Adult Intelligence Scale* (142), and *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (144).

N. GENERAL SUMMARY

The evidence indicates that at least Progressive Matrices (1938) has been very widely used, especially in England, clinically and in the British Services, but that it could and should be improved, with respect to both reliability and validity, especially for use in important diagnostic decisions about an individual. It is not a substitute in any sense for the Binet or

Wechsler tests, nor for any verbal or nonverbal group test of mental ability, but is perhaps an almost equally useful supplement, and shows intercorrelations with such tests perhaps as high as they show with one another. The evidence is not unequivocal that it is largely a pure test of Spearman's construct, *g*, even if such a device were possible. It has nevertheless been a useful research tool, especially in the study of the growth and deterioration of mental efficiency. It has especial value for use with special groups, such as the deaf and spastics. Its value for cross-cultural studies needs to be explored further. The revisions of *PM* (1938), *PM* (1947), Sets *A*, *Ab*, *B*, and *PM* (1947), Sets I and II are still in need of cross-validation research.

REFERENCES

1. ADKINS, D., & LYERLY, S. B. Factor Analysis of Reasoning Tests. Chapel Hill, N. C.: Univ. North Carolina Press, 1952.
2. ADOCK, C. A re-analysis of Slater's spatial judgment research. *Occup. Psychol.* ,1948, **22**, 213-216.
3. ALON-BAKALIAR, SH. Matritest progresiviyet b' mivkon bashvasti. (Progressive Matrices in comparative testing.) *Hakinakh*, 1950-52, **24**, 156-159. Seen in abstract. *Psychol. Abstr.*, 1953, **27**, 8035.
4. BANKS, C. Factor analysis applied to current psychological problems with special reference to data from *HM* Forces. (Unpublished doctor's thesis, Univ. London, 1945.)
5. ———. Factor analysis of assessments for Army recruits. *Brit. J. Psychol., Statist. Sect.*, 1949, **2**, 76-89.
6. ———. Progressive Matrices (1947), Sets *A*, *Ab*, *B*. A Review in O. K. Buros (*Ed.*), *Fourth Mental Measurements Yearbook*. Highland Park, New Jersey: Gryphon Press, 1953.
7. BANKS, C., & SINHA, U. An item-analysis of the Progressive Matrices Test and Binet. *Brit. J. Psychol., Statist. Sect.*, 1951, **4**, 91-94.
8. BARRATT, E. S. The relationship of the Progressive Matrices (1938) and the Columbia Maturity Scale to the *WISC*. *J. Consult. Psychol.*, 1956, **20**, 294-296.
9. BERNYER, G. Analyse factorielle de tests d'aptitude technique. *Année Psychol.*, 1946-47, **47-48**, 198-211.
10. BOLIN, B. J. A comparison of Raven's Progressive Matrices (1938) with the *ACE* Psychological Examination and the Otis Gamma Mental Ability Test. *J. Consult. Psychol.*, 1955, **19**, 400.
11. BOLTON, F. B. Experiments with the Raven's Progressive Matrices—1938. *J. Educ. Res.*, 1955, **48**, 629-633.
12. BRADFORD, E. J. G. Performance tests in the diagnosis of mental deficiency. *Brit. J. Med. Psychol.*, 1943, **19**, 394-414.
13. BROMLEY, D. B. Primitive forms of response to the matrices test. *J. Ment. Sci.*, 1953, **99**, 374-393.
14. BURT, C. The differentiation of intellectual ability. *Brit. J. Educ. Psychol.*, 1954, **24**, 76-90.
15. CASINI-NENCINI, M. G. Ricerche sul Progressive Matrices sperimentato in soggetti normali ed ammalati psichiatrici. (Unpublished thesis, Univ. of Rome, 1952-53.)

16. Cassel, R. H. Qualitative evaluation of the Progressive Matrices Tests. *Educ. & Psychol. Measmt.*, 1949, **9**, 233-241.

17. Crown, S. Notes on an experimental study of intellectual deterioration. *Brit. Med. J.*, 1949, **2**, 684-685.

18. Davidson, M. Studies in the application of mental tests to psychotic patients. *Brit. J. Med. Psychol.*, 1939, **18**, 44-52.

19. Delys, L. Normes du test Matrix 1938 à l'armée Belge de 1946 à 1948. *Rev. Belg. Psychol. Pedag.*, 1953, **15**, 13-17.

20. Desai, M. M. The test-retest reliability of the Progressive Matrices Test. *Brit. J. Med. Psychol.*, 1952, **25**, 48-53.

21. ———. The relationship of the Wechsler-Bellevue Verbal Scale and the Progressive Matrices Test. *J. Consult. Psychol.*, 1955, **19**, 60.

22. Edholm, O. G., & Gibson, Q. H. Examination results and an intelligence test. *Lancet*, 1944, **247**, 294-296.

23. Emmett, W. G. Evidence of a space factor at 11+ and earlier. *Brit. J. Psychol., Statist. Sect.*, 1949, **2**, 3-16.

24. Esher, F. J. S., Raven, J. C., & Earl, C. J. C. Discussion on testing intellectual capacity in adults. *Proc. Roy. Soc. Med.*, 1942, **35**, 779-785.

25. Eysenck, H. J. Neurosis and intelligence. *Lancet*, 1943, **245**, 362-363.

26. ———. The effect of incentives on neurotics and the variability of neurotics as compared with normals. *Brit. J. Med. Psychol.*, 1944, **20**, 100-103.

27. ———. Types of personality—a factorial study of 700 neurotics. *J. Ment. Sci.*, 1944, **90**, 851-861.

28. Eysenck, H. J., & Halstead, H. The memory functions: I. A factorial study of 15 clinical tests. *Amer. J. Psychiat.*, 1945, **102**, 174-180.

29. Eysenck, M. D. An exploratory study of mental organization in senility. *J. Neurol. Neurosurg. Psychiat.*, 1945, **8**, 15-21.

30. ———. A study of certain qualitative aspects of problem solving behaviour in senile dementia patients. *J. Ment. Sci.*, 1945, **91**, 337-345.

31. ———. The psychological aspects of ageing and senility. *J. Ment. Sci.*, 1946, **92**, 171-181.

32. ———. Cognitive factors in epilepsy. *J. Neurol. Neurosurg. Psychiat.*, 1952, **15**, 39-44.

33. Foulds, G. A. Variations in the intellectual activities of adults. *Amer. J. Psychol.*, 1949, **62**, 238-246.

34. Foulds, G. A., & Raven, J. C. Intellectual ability and occupational grade. *Occup. Psychol.*, 1948, **22**, 197-203.

35. ———. Normal changes in the mental abilities of adults as age advances. *J. Ment. Sci.*, 1948, **94**, 133-142.

36. ———. An experimental survey with Progressive Matrices (1947), Sets I and II. *Brit. J. Educ. Psychol.*, 1950, **20**, 104-110.

37. Gabriel, K. R. The simplex structure of the Progressive Matrices Test. *Brit. J. Statist. Psychol.*, 1954, **7**, 9-14.

38. Germain, J. Ensayo de un test perceptivo de inteligencia (Raven test). *Rev. Psicol. Gen. & Appl., Madrid*, 1947, **2**, 255-264.

39. Gibson, Q. H. Intelligence tests and university careers of medical students. *Lancet*, 1948, **255**, 323-324.

40. Green, M. W., & Ewert, J. Normative data on Progressive Matrices (1947). *J. Consult. Psychol.*, 1955, **19**, 139-142.

41. Guilford, J. P. The structure of intellect. *Psychol. Bull.*, 1956, **53**, 267-293.

42. HALSTEAD, H. An analysis of the Matrix (Progressive Matrices) test results on 700 neurotic (military) subjects and a comparison with Shipley Vocabulary Test. *J. Ment. Sci.*, 1943, **89**, 202-215.

43. ———. Abilities of male mental hospital patients. *J. Ment. Sci.*, 1950, **96**, 726-733.

44. HALSTEAD, H., & SLATER, P. An experiment in the vocational adjustments of neurotic patients. *J. Ment. Sci.*, 1946, **92**, 509-515.

45. HARDING, D. W. Prognostic tests for students of architecture. *Occup. Psychol.*, 1943, **17**, 126-131.

46. HEIM, A., & TIMPANY, N. Letter to the editor. *Lancet*, 1944, **294**, 392.

47. HEIM, A., & WALLACE, J. The effects of repeatedly retesting the same group on the same intelligence test. *Quart. J. Exp. Psychol.*, 1950, **2**, 19-32.

48. HIMMELWEIT, H. T. The intelligence-vocabulary ratio as a measure of temperament. *J. Personal.*, 1945, **14**, 93-105.

49. ———. Speed and accuracy of work as related to temperament. *Brit. J. Psychol.*, 1946, **36**, 132-143.

50. HOLDEN, R. H. Improved methods in testing cerebral palsied children. *Amer. J. Ment. Defic.*, 1951, **56**, 349-353.

51. HOULISTON, M. Selection tests given to nursing applicants. *Nurs. Times*, 1946, **42**, 808-810.

52. ———. Note on selection tests given to nursing applicants. *Nurs. Times*, 1947, **43**, 331.

53. HUNT, W. A., & FRENCH, E. The Navy-Northwestern Matrices Test. *J. Clin. Psychol.*, 1952, **8**, 65-74.

54. JASTAK, J. Problems of psychometric scatter analysis. *Psychol. Bull.*, 1949, **46**, 177-197.

55. JOHNSON, E. Sex differences and variability in the performance of retarded children on Raven, Binet, and Arthur tests. *J. Clin. Psychol.*, 1952, **8**, 298-301.

56. ———. Individual patterns of emotional functioning in children of comparable *IQ*'s—implications for education. *Amer. J. Ment. Defic.*, 1953, **58**, 681-686.

57. ———. The clinical use of Raven's Progressive Matrices to appraise potential for progress in play therapy: A study of institutionalized mentally and educationally retarded children. *Amer. J. Orthopsychiat.*, 1953, **23**, 391-405.

58. JONES, H. G. Comments on "The Validity and interchangeability of Terman-Merrill and Matrices test data" by D. Walton. *Brit. J. Educ. Psychol.*, 1956, **26**, 141.

59. KEATING, T. J. Progressive Matrices (1938). A Review in O. K. Buros *(Ed.), Third Mental Measurements Yearbook.* New Brunswick, N. J.: Rutgers Univ. Press, 1949.

60. KEIR, G. The Progressive Matrices as applied to school children. *Brit. J. Psychol., Statist. Sect.*, 1949, **2**, 140-150.

61. KLONOFF, H. An exploratory study and analysis of the Wechsler-Bellevue Intelligence Scale and the Raven Progressive Matrices. (Unpublished master's thesis, Univ. Toronto, 1951.)

62. LEVINE, B., & ISCOE, I. A comparison of Raven's Progressive Matrices with a short form of the Wechsler-Bellevue. *J. Consult. Psychol.*, 1954, **18**, 10.

63. ———. The Progressive Matrices (1938), the Chicago Non-Verbal, and the *W-B* on an adolescent deaf population. *J. Clin. Psychol.*, 1955, **11**, 307-308.

64. LEYBOURNE-WHITE, G. *Report on Capacity and Opportunity for University Education.* Manchester, England: Univ. Manchester Press, 1946.

65. ———. Opportunity and capacity for university education. *Pilot Papers,* 1947, **2**, 69-83.

66. Line, W. The growth of visual perception in children. *Brit. J. Psychol., Monog. Suppl.,* 1931, No. 5.

67. Looze, J. Essai de selection pre-universitaire. (Study of university entrance.) *Travail Hum.,* 1954, **17**, 224-234.

68. Martin, A. W., & Wiechers, J. E. Raven's Colored Progressive Matrices and the Wechsler Intelligence Scale for Children. *J. Consult. Psychol.,* 1954, **18**, 143-144.

69. Miller, F., & Raven, J. C. The influence of positional factors on the choice of answers to perceptual intelligence tests. *Brit. J. Med. Psychol.,* 1939, **18**, 35-39.

70. Misiti, R. Presentazione del reattivo "Progressive Matrices" (adulti e fanciulli) di J. C. Raven. *Ricerca Scient., Rome,* 1954, **24**, 1052-1058.

71. Moore, B. G. R., & Peel, E. A. Predicting aptitude for dentistry. *Occup. Psychol.,* 1951, **25**, 192-199.

72. Morali-Daninos, Thyerry-Mieg, Granien. Etallonage du test Matrix Progressive 1938 administrée à temps limité. Clamart: Editions scientifiques, 1950.

73. Notcutt, B. The distribution of scores on Raven's Progressive Matrices Test. *Brit. J. Psychol.,* 1949-50, **40**, 68-70.

74. ———. The measurement of Zulu intelligence. *J. Soc. Res., Pretoria,* 1950, **1**, 195-206.

75. O'Connor, N., & Tizard, J. Predicting the occupational adequacy of certified mental defectives. *Occup. Psychol.,* 1951, **25**, 205-211.

76. Oléron, P. Etude sur les capacités intellectuelles des sourds-muets. *Année psychol.,* 1946-47, **47-48**, 136-155.

77. Oléron, P. Tests non verbaux d'intelligence et réussite scolaire chez les élèves sourds-muets. *J. de Psychol.,* 1949, **42**, 331-338.

78. ———. A study of the intelligence of the deaf. *Amer. Ann. Deaf.,* 1950, **95**, 179-195.

79. Orton, R., & Martin, D. R. Psychiatric screening of medical students. *Lancet,* 1948, **255**, 321-323.

80. Oxlade, M. An experiment in the use of psychological tests in the selection of women trainee telephone mechanics. *Bull. Industr. Psychol. & Person. Prac.,* 1946, **2**, 26-32.

81. Penrose, L. S., & Raven, J. C. A new series of perceptual tests: Preliminary communication. *Brit. J. Med. Psychol.,* 1936, **16**, 97-104.

82. Raven, J. C. Mental tests used in genetic studies. (Unpublished master's thesis, Univ. London, 1936.)

83. ———. Progressive Matrices: A perceptual test of intelligence, 1938, Individual Form. London: Lewis, 1938.

84. ———Progressive Matrices: A perceptual test of intelligence, 1938, Sets *A, B, C, D,* and *E.* London: Lewis, 1938.

85. ———. Guide to Using Progressive Matrices (1938). London: Lewis, 1938-1954.

86. ———. The R. E. C. I. series of perceptual tests: An experimental survey. *Brit. J. Med. Psychol.,* 1939, **18**, 16-34.

87. ———. Matrix tests. *Ment. Hlth, London,* 1940, **1**, 10-18.

88. ———. Standardization of Progressive Matrices, 1938. *Brit. J. Med. Psychol.,* 1941, **19**, 137-150.

89. ———. Testing the mental ability of adults. *Lancet,* 1942, **242**, 115-117.

90. ———. Progressive Matrices (1947): Sets I and II. London: Lewis, 1947.

91. ———. Raven's intelligence test: Reply by John C. Raven. *Brit. Med. J.*, 1947, 4510, 872.

92. ———. A comparison of skill and awkwardness. (Privately printed by the Crichton Royal, Dumfries, Scotland, 1948.)

93. ———. A note on Patrick Slater's comment. *Brit. J. Psychol.*, 1948, **39**, 21.

94. ———. Progressive Matrices 1947, plan and use of the scale, with the report of an experimental survey by G. A. Foulds. (Privately printed by the Crichton Royal, Dumfries, Scotland, 1948.)

95. ———. The comparative assessment of intellectual ability. *Brit. J. Psychol.*, 1948, **39**, 12-19.

96. ———. Progressive Matrices (1947), Sets *A*, *Ab*, *B*: Board and Book Forms. London: Lewis, 1949.

97. ———. Guide to Using Progressive Matrices (1947), Sets *A*, *Ab*, *B*, *Revised*. London: Lewis, 1949-1951.

98. ———. Manual d'instruction Matrix 1948. Clamart: Editions Scientifiques et Psychotechniques, 1950.

99. Raven, J. C., & Waite, A. Experiments on physically and mentally defective children with perceptual tests. *Brit. J. Med. Psychol.*, 1939, **18**, 40-43.

100. Rimoldi, H. J. A. A note on Raven's Progressive Matrices Test. *Educ. & Psychol. Measmt.*, 1948, **8**, 347-352.

101. ———. Study of some factors related to intelligence. *Psychometrika*, 1943, **13**, 27-46.

102. Rimoldi, H. P. A., Cortada, N., & Velasco, E. Ensayo de tipificacion de una prueba mental (Progressive Matrices de Raven.) Standardizing use of a mental test. *Publ. Inst. Psicol. Exp., Univ. Cuyo*, 1945, **1**, 83-114.

103. Rimoldi, H. J. A., Velasco, E. S., de San Martin, R., & Bührer, L. Tipificacion de los "Progressive Matrices" de Raven. (Standardization of the Progressive Matrices of Raven.) *Publ. Inst. Psicol. exp., Univ. Cuyo*, 1947, **2**, 1-24.

104. Roberts, J. A. F. Further observations on the efficiency of the Progressive Matrices Test. (Report submitted to the War Cabinet Expert Committee on the work of psychologists and psychiatrists in the Services, 1943.)

105. Rudolf, G. de M. The Kent and other tests used on the same subjects. *J. Ment. Sci.*, 1948, **44**, 452-458.

106. Seeger, E. Dominante Fehler im Test "Progressive Matrices" bei grosser Intelligenz. (Predominant errors in the Progressive Matrices of individuals with superior intelligence.) *Diagnostica*, 1955, **1**, 30-32.

107. Sinha, U. A study of the reliability and validity of the Progressive Matrices. (Summary of thesis submitted for the Degree of M.A., University of London, 1950.) *Brit. J. Educ. Psychol.*, 1951, **21**, 238-239.

108. Slater, P. Scores of different types of neurotics on tests of intelligence. *Brit. J. Psychol.*, 1945, **35**, 40-42.

109. ———. The association between age and score in the Progressive Matrices Test. *Brit. J. Psychol., Statist. Sect.*, 1947, **1**, 64-69.

110. ———. Comment on "The comparative assessment of intellectual ability." *Brit. J. Psychol.*, 1948, **39**, 20-21.

111. Smith, D. I. A survey of the intelligence and attainments of a group of deaf children. (Unpublished thesis, Univ. Aberdeen, 1951.)

112. Smyth, D. H. Some principles in selection of medical students. *Brit. Med. J.*, 1946, **2**, 357-367.

113. Spearman, C. L'examen de l'intelligence. *Travail Hum.*, 1937, **5**, 385-391.

114. ———. Measurement of intelligence. *Scientia, Bologna*, 1938, **64**, 75-82.

115. ———. Intelligence tests. *Eugen. Rev.*, 1939, **30**, 249-254.

116. ———. Theory of a general factor. *Brit. J. Psychol.*, 1946, **36**, 117-131.

117. SPEARMAN, C., & JONES, L. L. W. Human Ability. London: Macmillan, 1950.

118. STACEY, C. L., & CARLETON, F. O. The relationship between Raven's Colored Progressive Matrices and two tests of general intelligence. *J. Clin. Psychol.*, 1955, **11**, 84-85.

119. STACEY, C. L., & GILL, M. R. The relationship between Raven's Colored Progressive Matrices and two tests of general intelligence for 172 subnormal adult subjects. *J. Clin. Psychol.*, 1955, **11**, 86-87.

120. STEPHENSON, W. Tetrad differences for non-verbal tests. *J. Educ. Psychol.*, 1931, **22**, 167-185.

121. TAIBL, R. An investigation of Raven's Progressive Matrices as a tool in the psychological evaluation of cerebral palsied children. (Unpublished doctor's dissertation, Univ. Nebraska, 1951.)

122. TIZARD, J., O'CONNOR, N., & CRAWFORD, J. M. The abilities of adolescent and adult high-grade male defectives. *J. Ment. Sci.*, 1950, **96**, 889-907.

123. TRACHT, V. S. Preliminary findings on testing the cerebral palsied with Raven's Progressive Matrices. *J. Except. Child.*, 1948, **15**, 77-79.

124. VERNON, P. E. The reliability and validity of the Progressive Matrices Test. (S. P. Test Follow-Up Report, No. 146, 1942.)

125. VERNON, P. E. Psychological tests in the Royal Navy, Army, and A. T. S. *Occup. Psychol.*, 1947, **21**, 53-74.

126. ———. Research on personnel selection in the Royal Navy and the British Army. *Amer. Psychol.*, 1947, **2**, 35-51.

127. ———. The variations of intelligence with occupation, age, and locality. *Brit. J. Psychol., Statist. Sect.*, 1947, **1**, 52-63.

128. ———. Occupational norms for the 20-minute Progressive Matrices Test. *Occup. Psychol.*, 1949, **23**, 58-59.

129. ———. The structure of practical abilities. *Occup. Psychol.*, 1949, **23**, 81-96.

130. ———. Recent developments in the measurement of intelligence and special abilities. *Brit. Med. Bull.*, 1949, **6**, 21-23.

131. ———. An application of factorial analysis to the study of test items. *Brit. J. Psychol., Statist. Sect.*, 1950, **3**, 1-15.

132. ———. The Structure of Human Abilities. London: Methuen, 1950.

133. VERNON, P. E., & PARRY, J. B. Personnel Selection in the British Forces. London: Univ. London Press, 1949.

134. VINCENT, D. V. The linear relationship between age and score of adults in intelligence tests. *Occup. Psychol.*, 1952, **26**, 243-249.

135. WALL, W. D. Progressive Matrices (1947), Sets *A*, *Ab*, *B*. A Review in O. K. Buros (*Ed.*), *Fourth Mental Measurements Yearbook*. Highland Park, N. J.: Gryphon Press, 1953.

136. WALTON, D. The validity and interchangeability of Terman-Merrill and Matrices test data. *Brit. J. Educ. Psychol.*, 1955, **25**, 190-194.

137. ———. Reply to H. Gwynne Jones' "Comments on the 'The validity and interchangeability of Terman-Merrill and Matrices test data.'" *Brit. J. Educ. Psychol.*, 1956, **26**, 142-143.

138. WARBURTON, F. W. The ability of the Gurka recruit. *Brit. J. Psychol.*, 1951, **42**, 123-133.

139. ———. Relationship between intelligence and size of family. *Eugen. Rev.*, 1951, **43**, 36-37.

140. ———. Relationship between the intelligence of students and size of family. *Eugen. Rev.*, 1952, **43**, 188.

141. WECHSLER, D. Progressive Matrices, Sets *A*, *B*, *C*, *D*, and *E*, 1938. A Review in O. K. Buros (*Ed.*), *Third Mental Measurements Yearbook.* New Brunswick, N. J.: Rutgers Univ. Press, 1949.

142. ———. Manual for the Wechsler Adult Intelligence Scale. New York: Psychol. Corp., 1955.

143. WESTBY, G. Progressive Matrices, Sets *A*, *B*, *C*, *D*, and *E*, 1938. A Review in O. K. Buros (*Ed.*), *Fourth Mental Measurements Yearbook.* Highland Park, N. J.: Gryphon Press, 1953.

144. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, **51** (Supplement).

Psychology Service
VA Hospital
East Orange, New Jersey

The Journal of Genetic Psychology, 1958, **93**, 229-236.

THE DEVELOPMENT OF PAW PREFERENCES IN CATS AND MONKEYS*

Department of Psychology, Stanford University

J. M. WARREN[1]

A. INTRODUCTION

Cole (1) recently determined the distribution of lateral preferences in a group of 60 cats, and found that 58 per cent of the animals used one paw on more than 75 per cent of the trials, and that 42 per cent were ambidextrous. He concluded that an innate anatomical characteristic, asymmetrical distribution of pyramidal tract fibers, is probably more significant than environmental factors in the genesis of lateral preferences.

Now, the fact that one half of the spinal cord receives a larger number of crossed and uncrossed cortico-spinal fibers establishes the possibility of an innate anatomic determination of laterality, but Cole's data were inadequate to establish a necessary connection between lateral preferences and their presumed structural basis, since it was not shown that individual cats with, say, strong left preferences had a larger number of cortico-spinal fibers in the left half of the spinal cord. Nor, since he only tested his animals once, did Cole collect sufficient data to confirm the behavioral implications of his anatomic hypothesis: (*a*) that individual tests of lateral preference would yield very high retest correlations, and (*b*) that different tests of paw preference would yield very high intertest correlations.

The purpose of this paper is to describe some observations on the development of lateral preferences in cats and monkeys over several months of continuous practice, which permit more rigorous testing of the behavioral corollaries of Cole's anatomic theory of laterality.

B. METHOD

1. *Subjects*

A total of 86 cats was studied. Detailed results, however, will be presented only for a group of 31 cats which were treated most similarly to the 17 monkeys studied. The monkeys and the comparable group of cats

---

*Received in the Editorial Office on September 23, 1957.

[1] This investigation was supported (in part) by Research Grants M-835 and M-1364 from the National Institute of Mental Health, U. S. Public Health Service. Some of the data were collected at the University of Oregon.

were first tested as naïve *S*s and were retested twice over a period of approximately a year.

2. *Apparatus*

The Wisconsin General Test Apparatus (2, Figure 1) was used throughout the experiment.

3. *Procedure*

The cats were tested on two tests (I and III) and the monkeys on all of the following three tests:

I. *S* was required only to pick up pieces of food from the test tray of the W.G.T.A. This task has already been described in detail (7).

II. The monkey was obliged to displace a thin cardboard square to uncover a piece of food placed in one of the foodwells in the test tray before picking up the food.

III. *S* was required to uncover food by removing a wooden block.

The dependent variable in all tests was the paw which the animal used in carrying the food to his mouth. All of the tests consisted of at least 100 trials, and the position of the food was randomly varied between trials to control for possible positional biases.

The first preference tests (Series *A*) were administered during the taming of the animals and were given on the first days that the *S*s were placed in the *WGTA*. Thus the cats and monkeys were both experientially and experimentally naïve with respect to the test situation. Daily testing in the *WGTA* on learning problems which involved continuous practice in conveying small pieces of food from test tray to mouth intervened between the first tests and the second (Series *B*) and third (Series *C*) cycles of testing. The data on paw preference were, in fact, collected incidentally to learning experiments, the scheduling of which dictated slight differences in the inter-test intervals for the individual cats and monkeys.

The naïve cats were tested on Tests I and III during their first week of work in the laboratory (Series *A*); a median interval of six months intervened between Series *A* and *B*. The average time between Series *B* and *C* was five months.

For the monkeys, three months intervened between Tests I and II during the first cycle of testing (*A*), and 10 weeks between Tests II and III. All three tests were given within the same week during the second and third series (*B* and *C*), and approximately three months elapsed from the last test in Series *A* to Series *B*, and also between *B* and *C*. Thus the longitudinal study of the cats and monkeys covered a period of about 11 months.

C. RESULTS

1. *The Distribution of Lateral Preferences*

The percentage of *S*s which used one paw or the other on 80 per cent or more of the trials on each repetition of the tests is shown in Table 1.

TABLE 1
PERCENTAGE OF SUBJECTS WITH GREATER THAN 80 PER CENT PREFERENCE ON SUCCESSIVE ADMINISTRATIONS OF TESTS

	17 Monkeys				31 Cats		
	A	B	C		A	B	C
I	30	83	71	I	40	61	81
II	59	83	100	II	—	—	—
III	53	89	100	III	42	80	90

These data indicate that the proportion of animals showing a significant preference increases by about 40 per cent over the period of several months; this trend is generally consistent over all tests, for both cats and monkeys. It should be noted, too, that for both cats and monkeys a higher proportion of animals manifest preferences on the test(s) which require more complex responses than simply securing a piece of food lying free on the test tray (I).

2. *Reliability of the Tests*

(*a*). *Split-half reliability*. Product-moment correlations indicating the relation between the first and second half scores on each test are given in Table 2. The *r*'s indicate that the tests yield highly consistent results; on

TABLE 2
SPLIT HALF RELIABILITY FOR HANDEDNESS TESTS AT SUCCESSIVE STAGES OF TRAINING

	17 Monkeys Series			31 Cats Series		
Test	A	B	C	A	B	C
I	.65	.97	.90	.89	.99	.97
II	.80	.94	.97	—	—	—
III	.76	.97	.99	.94	.97	.99

Series *B* and *C* all of the values are greater than .90. The correlations are, particularly for the monkeys, lower on Series *A*, which suggests that some learning or adaptation is required for the animals' responses to become highly consistent. A slightly higher degree of consistency characterizes the results for the cats compared to the monkeys.

(*b*). *Retest reliability*. The correlations between successive administrations of the tests are presented in Table 3. A marked increase in retest consistency is observed in both species, and for both cats and monkeys one notes that the more complex Test III is more reliable than I.

TABLE 3
RETEST RELIABILITY OF HANDEDNESS TESTS FOR CATS AND MONKEYS

	17 Monkeys		31 Cats	
	Series Correlated		Series Correlated	
Test	A&B	B&C	A&B	B&C
I	.16	.81	.59	.87
II	.24	.79	—	—
III	.89	.99	.72	.95

3. *The Generality of Paw Preferences*

The generality of paw preferences can be estimated quantitatively by the correlations between the different tests. The intertest correlations at the three stages of training are presented in Table 4, which shows a consistent increase in the magnitude of the correlations from the earlier to later series. This increase is found for all tests and for both species.

TABLE 4
CORRELATIONS BETWEEN HANDEDNESS TESTS AT SUCCESSIVE STAGES OF TRAINING

	17 Monkeys			31 Cats		
	Series			Series		
Tests	A	B	C	A	B	C
I & II	.01	.69	.80	—	—	—
I & III	.29	.91	.91	.45	.91	.94
II & III	.54	.79	.94	—	—	—

D. DISCUSSION

Comparison of the results of this experiment with the findings obtained in similar investigations of the incidence of paw preferences reveals quite satisfactory correspondence. Cole found that 58 per cent of his unrestrained naïve cats used one paw in 75 per cent or more of the trials when required to remove a piece of meat from a glass tube (1, Figure 1). In spite of the fact that our cats were confined in the *WGTA* and drew in freely accessible pieces of meat, we found that 35 out of 66, 53 per cent, of the naïve cats used the same paw on 75 per cent or more of the trials when first tested on Test I. Warren (7) determined the proportion of preferences in a sample of 84 experimentally sophisticated monkeys on Test I; a comparison of the results of this earlier study with those obtained from the present group of monkeys on Test I, *A* is presented in Table 5. The slightly higher percentage of animals with preferences in the present experiment probably reflects the fact that the *S*s in this study were a more homogeneous group with respect to duration of training in the *WGTA*.

Having established, insofar as possible, the validity of the present results, let us consider the implications of the findings that both cats and monkeys show a progressive increase in split-half and retest reliabilities, in intertest consistency, and in the proportion of *S*s manifesting significant lateral preferences on successive series of tests.

TABLE 5

PERCENTAGE OF *S*S WITH LATERAL PREFERENCES IN FOUR MAMMALIAN SPECIES

	Rat	Cat	Monkey		Chimpanzee
Preferred member used > 90 per cent	*	71	55	59	46
Preferred member used > 80 per cent	±85**	84	63	70	64
Number of subjects	105	51	84	17	31
Source	(6)		(7)		(2)

*Not given.
**Number using one hand > 75 per cent.

It must be concluded that environmental factors are very influential in the determination of lateral preferences. The data permit identification of two such factors: (*a*) an experiential or practice effect, and (*b*) a task variable. The evidence for the importance of practice has been summarized in the preceding paragraph. All three of the tests require the same final motor response, carrying or dragging the food from test tray to *S*'s mouth, yet we find that the apparently trivial alteration of the test situation associated with covering the food with a piece of cardboard (II) or a wooden block (III) produces a marked change in the results obtained. A larger proportion of animals, both cats and monkeys, have preferences on the more complex Test III than on the simpler Test I; indeed, 29 per cent of the monkeys expressed no significant preference on Test I, *C*, but the percentage without preference on Tests II, *C*, and III, *C*, was zero.

The results presented in this paper cannot be construed as a refutation of Cole's hypothesis of an anatomic basis of lateral preferences. On the contrary, the findings reported here represent the most convincing evidence in favor of such an explanation. In Kounin's (5) methodologically sophisticated study of handedness in monkeys, he found that the manual performance of his *S*s was inconsistent on repetitions of the same test and between different tests. Similar results were obtained in the earliest stage (*A*) of the present investigation, but it was found that after months of training in the apparatus both cats' and monkeys' scores approached almost perfect consistency and generality, as measured by split-halves and intertest correlations.

It should be concluded, then, that Cole's explanation of laterality is more naïve than wrong. It is patently as impossible to answer the heredity-environment pseudoproblem with respect to lateral preferences as with regard to intelligence. Both anatomic and environmental factors are necessarily and importantly involved in any test performance. If animals are tested on the same tasks for a sufficiently long period of time they develop highly consistent responses in the practiced tasks, and the individual differences in performance may reflect variations in cortico-spinal innervations as suggested by Cole.

Moreover, it should be pointed out that there is considerable evidence to suggest that the rôle of learning and the importance of environmental factors becomes progressively more important in determining laterality in the higher mammalian orders. Our monkeys gave lower reliability retest and intertest correlations in the first stages of testing than did the cats, but yielded almost identical values on the later stages of testing. This finding suggests less rigid determination of response in the monkeys. Greater variability of response to different environmental conditions in primates than in carnivores is obvious when we compare the performance of the cats and monkeys on Tests I and III. On Test I, the reliability coefficients and proportion of animals with marked preferences is higher for the cats than for the monkeys; the relationship is completely reversed on Test III, on which the monkeys have stronger and more consistent preferences. This finding is compatible with Kounin's observation that complex tests elicit more consistent preferential responses from monkeys than do simpler tasks.

The suggestion that the phyletic trend is toward greater variability in manual behavior and decreasing anatomic determination is supported by a consideration of the data obtained in several experimental situations essentially similar to our Test I. Data have been obtained from naïve rats (6) and sophisticated cats, monkeys (7), and chimpanzees (2), and are summarized in Table 5, which shows the percentage of *S*s who used one paw on more than 80 per cent and more than 90 per cent of the trials. (The percentages for the cats are based upon the 31 *S*s most comparable to the monkeys plus the data obtained from 20 sophisticated animals which were tested on Test I after about eight months of training in the *WGTA*). It is readily apparent that there is a consistent decrease in the percentage of animals showing marked preferences from rat to chimpanzee. The magnitude of this effect is somewhat attenuated by the fact that naïve rats are compared with sophisticated representatives of the other species, but we can estimate the effect of this defect by noting that, compared to the 85 per cent of naïve rats, 52 per cent of the naïve cats and 35 per cent of the naïve monkeys preferred one hand over the other on 75 per cent of the trials on Test I, *A*. Thus, the evolutionary trend appears to be from relatively strong initial lateral preferences in the definitely quadrupedal species toward initial ambilaterality in the primates, including man. Hildreth's (4) review indicates that if there were no cultural bias, right- and left-handedness would be almost equally frequent among humans. This conclusion is based upon the facts that very little lateral preference is found in human infants, in adults in primitive societies, or in adults in complex societies when engaged in manual tasks which are not socially significant or culturally defined.

E. SUMMARY

The development of lateral preferences on two different tests by 31 cats and 17 monkeys was compared over approximately 11 months by determining preferences in the experimentally naïve *S*s during their first adaptation to the Wisconsin General Test Apparatus and upon two subsequent occasions. During the intervals between the three repetitions of the preference tests, the *S*s were tested daily upon learning problems in the *WGTA*, which required the same manual response as the preference tests. (The monkeys were studied on a third handedness test which could not be administered to the cats.) The following results were obtained:

1. For both cats and monkeys, split-half and retest reliability correlations, intertest correlations and the proportion of *S*s manifesting significant preferences increased progressively from the first to the third administration of the tests.
2. The proportion of monkeys with significant preferences was smaller than the proportion of cats on the simpler test, but higher on the more complex task on the second and third cycles of testing.

It was concluded that learning and environmental (task) variables are important in the development of lateral preferences. Consideration of the results obtained from rats, chimpanzees, and man suggests the possibility that the importance of environmental determinants, relative to anatomical factors, increases with phyletic status within the mammalian series.

REFERENCES

1. COLE, J. Paw preference in cats related to hand preference in animals and men. *J. Comp. & Physiol. Psychol.*, 1955, **48**, 137-140.
2. FINCH, G. Chimpanzee handedness. *Science*, 1941, **94**, 117-118.

3. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51-65.
4. HILDRETH, G. The development and training of hand dominance. *J. Genet. Psychol.*, 1949, **75**, 197-275; 1950, **76**, 39-144.
5. KOUNIN, J. S. Laterality in monkeys. *J. Genet. Psychol.*, 1938, **52**, 375-393.
6. TSAI, L., & MAURER, S. Right handedness in white rats. *Science*, 1930, **72**, 436-438.
7. WARREN, J. M. Handedness in the rhesus monkey. *Science*, 1953, **118**, 622-623.

Department of Psychology
Stanford University
Stanford, California

The Journal of Genetic Psychology, 1958, **93**, 237-248.

REEVALUATION OF THE REVISED STANFORD-BINET SCALE, FORM *L* VOCABULARY AS A TEST OF INTELLIGENCE FOR THE KINDERGARTEN AND PRIMARY SCHOOL CHILD*

Psychological Center, Graduate School of Education, Yeshiva University

BORIS M. LEVINSON[1]

A. INTRODUCTION

There has been a radical change in American life over the past 25 years. We have lived through a depression, a war, and a readjustment period. We have new media of mass education: talking pictures, television, etc. The level of education of American children has been immeasurably improved (22). More children go to school and to college (22). Our standard of living has sky-rocketed (21). In a word, our children have different experiences and live in a different culture than we did. The mean scores on certain intelligence tests have also changed. The American today achieves a much higher score on an intelligence test designed for his father, than his father did (20). What implications should we draw from this? Has the intelligence of the American people gone up for the last 30 years and hence the better performance? Is the higher rating a cultural artifact? If so, what does an *IQ* on such a test as the Revised Stanford-Binet (18) mean in terms of ability, school achievement, etc.?

We become even more perplexed when we learn about the relationship existing between the Binet and the more recent Wechsler Intelligence Scale for Children (23). Why does the *WISC* almost always give a lower *IQ* than the Binet (1, 3, 4, 8, 10, 13, 19, 24)? After all, the *WISC*, too, has been standardized on a representative sample of American children. Is the Binet too easy?

We hypothesized that the change in culture made certain Binet tasks more familiar than others. In other words, there has been a differential change in the difficulty of test items, some of them becoming easier and others more difficult. To test out this hypothesis, we decided to make a study of the relation of the Binet vocabulary to the rest of the scale. Previous research has shown that the vocabulary may be used as an individual test of intelligence

---

*Received in the Editorial Office on October 8, 1957.

[1] Based on a paper presented at the Annual Convention of the American Psychological Association, New York, August 31st, 1957.

(6, 12) and there are studies available comparing the Binet *MA* with the vocabulary (6, 12).

We do not go along with the thesis that a group's adjustment will affect its vocabulary (15). In our experience, we found that hysterics do not necessarily have a low vocabulary nor do obsessive compulsives have a high vocabulary. The cultural level will tell in the final analysis. We believe that the vocabulary is a test indicating information one has acquired in his cultural milieu and that it directly reflects the subculture of the home. We have seen dull children, coming from superior homes, who have had extensive vocabularies, and very bright children, coming from impoverished homes, with limited vocabularies.

B. PROBLEMS

The problems to which we addressed ourselves were as follows:

1. Has the Binet vocabulary retained its relative contribution as an index of intelligence as compared with the full scale Binet on the kindergarten and primary school level?
2. What influence does a different sub-culture (foreign birth) have on the Binet vocabulary, as compared to the full scale Binet?
3. Has there been a displacement in the order of difficulty of the Binet vocabulary?

In view of the fact that the population available to us was on the kindergarten and primary school level, our analysis of the Binet vocabulary was in terms of this population.

C. PREVIOUS RESEARCH

We find conflicting results dependent upon the sample studied as to the relationship existing between Binet vocabulary and Binet *MA*.

Terman (18) considered the vocabulary to be the most valuable single test of the scale. He based this conclusion on the correlation between the vocabulary scores and mental ages, this being on the average of .81.

McNemar (12), who made an analysis of the Binet standardization data, mentioned that the vocabulary score gives the correlation of .71 with a composite *MA* at age eight ($N = 200$). He prepared norms for "the means and standard deviations for vocabulary test, scored as number of words passed," for *CA* 7 to 18.

Recently Cureton (5) reviewed McNemar's (12) statistical study of the Revised Stanford-Binet and calculated on the basis of McNemar's data mental age equivalents for the Revised Stanford-Binet vocabulary from five up to 40 words.

Elwood (6) made one of the earliest investigations of the Binet vocabulary. She made a study of 1,161 Pittsburgh school children who were tested during the year 1937-38 and whose mental age reached or exceeded four. She found the correlation between vocabulary score and *MA* to be .978. The mean vocabulary for *MA* four ($N = 56$) was 1.57; for *MA* five ($N = 97$), 3.49; for *MA* six ($N = 190$), 4.82; for *MA* seven ($N = 175$), 6.79; for *MA* eight ($N = 158$), 7.92; and *MA* nine ($N = 111$), 9.18. Elwood felt that "the vocabulary norms for the ages studied were approximately correct for the children used in this study."

However, the correlations found by Spache (16) were not as high. For a small sample, he found the correlation between the vocabulary score and *MA* of .847 ($N = 65$) and vocabulary *MA* and *MA* Binet of .915 ($N = 66$). In this population, an underestimation of scale *MA* by approximately one year or more occurred in 14 per cent and overestimation in 40 per cent. However, in a larger population ($N = 86$), overestimation of scale *MA* by more than one year occurred in 9 per cent, an underestimation in only 15 per cent of the cases.

A review of Feifel and Lorge's (7) study indicated that the vocabulary *MA* of their group was somewhat below McNemar's (12) and Cureton's (5) norms. They studied samples of 100 native born white children with no language handicaps, at four different age levels. Their findings were as follows: Children with an *MA* of 80.1 and an *IQ* of 100.3 had a vocabulary score of 6.2. Those with an *MA* of 91.1 and an *IQ* of 102.3, had a vocabulary score of 7.3, and those with an *MA* of 104.2 and an *IQ* of 103 had a vocabulary score of 8.3. Those with an *MA* of 117.4 and an *IQ* of 104.1 had a vocabulary score of 9.9.

D. PROCEDURE

Seven hundred and seventy-six consecutive records of children (whose test results were considered valid) were taken out of the files of the Psychological Center of Yeshiva University.

Six hundred and seventy (86.34 per cent) of these were children who applied for admission to day schools (Jewish religious schools) and who came in for routine determination of mental level. Six hundred and forty of our sample were native born and 136 foreign born. (We excluded from our sample children who were born either in England or Canada.) About two-thirds of the sample were boys. In our analysis of data, we paid no attention to sex differences, as previous research indicated that no statistically significant differences exist in the intelligence of the boys and girls at the preschool and first grade levels, to which the overwhelming majority of these children belonged (11). We had no data to indicate extent of bilingualism in our sample, nor did we have information as to parental occupation.

In the scoring of the vocabulary test, the standards laid down by Terman (18) and Pintner (14) were followed. No attempt was made at the qualitative analysis of the answers given to the vocabulary items. This undoubtedly would be a worth-while study in itself as it might reflect the cultural change in the meaning of words. Credit was given for any definition which met the criteria laid down by Terman and Pintner.

The Revised Stanford-Binet protocols were placed in three age groups and analyzed accordingly. These groups were as follows: Group *A*—*CA* 4 years to 5 years, 11 months; Group *B*—*CA* 6 years to 7 years, 11 months; Group *C*—*CA* 8 years to 9 years, 11 months.

E. RESULTS

Table 1 presents the background data of the children used in this study. Note that both the native and foreign born children were of above average intelligence and that 518 (66.75 per cent) subjects were below six years of age.

TABLE 1

MEANS AND *SD*s IN *CA*, *MA*, AND *IQ* FOR NATIVE AND FOREIGN BORN FOR *CA*s 4 TO 5-11, 6 TO 7-11, 8 TO 9-11

	CA		*MA*		*IQ*	
	Native	Foreign	Native	Foreign	Native	Foreign
			4 to 5-11			
N	446	72	446	72	446	72
Mean	63.58	63.75	72.00	67.79	113.24	106.34
SD	5.76	3.80	10.92	9.72	14.60	13.20
			6 to 7-11			
N	155	57	155	57	155	57
Mean	78.54	76.78	94.00	84.6	119.68	110.18
SD	8.28	6.42	16.68	10.20	16.05	16.99
			8 to 9-11			
N	39	7	39	7	39	7
Mean	103.71	106.07	129.1	129.8	124.48	122.37
SD	6.27	6.56	23.46	3.9	15.49	14.10

Table 2 presents the correlations between Binet *MA* and Binet vocabulary. It is to be noted that Binet *MA* equivalents are not available for Binet vocabulary of less than five words (5). Since 13.76 per cent of native born children and 49.26 per cent of foreign born children had a vocabulary score of less than five words (see Table 4), it was necessary to correlate the raw vocabulary scores with the Binet *MA*s. The correlations presented were in part spurious since the vocabulary test was included in the Binet *MA* determination. This spuriousness, however, did not affect our results too seriously, since, for our group, the vocabulary was credited, in the vast majority of cases, only on the sixth and eighth year levels. Even though the correlations were somewhat inflated, they still were lower than those reported in the literature. If confirmed by further research, this would indicate the inadvisability of the use of the Binet vocabulary alone as a short test of intelligence. However, if the full scale Binet were to be accepted as a valid criterion, the vocabulary was still a valid index of intelligence for the ages under consideration.

TABLE 2
CORRELATIONS BINET *MA* AND BINET VOCABULARY

	4 to 5-11		6 to 7-11		8 to 9-11	
	Native	Foreign	Native	Foreign	Native	Foreign
N	446	72	155	57	39	7
r	.62	.64	.70	.44	.70	.61

TABLE 3
THE MEANS OF THE MENTAL AGE EQUIVALENTS OF THE VOCABULARY OF THE NATIVE BORN CHILDREN AND STANDARDIZATION GROUP FOR SEVEN TO ELEVEN WORDS

	Native born children				*Standardization group*		
Words	Mean *MA*	*SD* *MA*	*N*	Words	Mean *MA*	*SD* *MA*	*N*
7	6.61	.787	102	7*	7.2**	1.00**	200
8	7.24	.728	77	8***	8.0***	1.2****	200
9	7.39	1.17	55	9***	8.8***	1.3****	200
10	7.58	1.29	53	10*	9.4**	1.4**	200
11	8.83	1.71	32	11***	10.2***	1.6****	200

*McNemar (12) Table 44, p. 41.
**McNemar (12) Tables 6 and 7, pp. 32-33.
***Cureton (5).
****Estimated from McNemar (12) Tables 6 and 7, pp. 32-33.

Table 3 shows the means and mental age equivalents of the vocabulary words from seven to 11 for native born children and for the standardization group. The means for the lower ages of the native born children were not reliable since the *CA* was not distributed normally and the average *MA*, even for the age group four to five-eleven, was six, bringing about a spuriously high *MA* vocabulary score. Neither were the means for the older ages reliable as there were too few cases at these ages.

We computed data from Table 3 and found that the combined mean *MA* ($N = 319$) for the native born children was 7.28 and combined *SD* (9,

p. 57) was 1.24. The combined mean *MA* ($N = 1{,}000$) for the standardization group was 8.72 and combined *SD* was 1.68. The difference between the means was 1.44, the *SE* of the difference was 0.086, and the critical ratio was 16.74. The difference was at .01 level of confidence. This indicated that the mean vocabulary *MA* of native born children overestimated the full Binet *MA* for native born children.

We computed data from Table 1 and found that the combined mean *MA* for the foreign born group ($N = 136$) was 6.85 with a combined *SD* (9, p. 57) of 1.50. The mean vocabulary was 5.03. We compared this with Terman's standardization group for five vocabulary words (5). The mean *MA* ($N = 200$) for this group as given by Cureton (5) is 5.33. *SD* estimated from McNemar (12 Table 7, p. 33) was .6. The difference between the means was 1.52, the *SE* of this difference was 0.13 and the critical ratio was 11.69. This is at .01 level of confidence. This indicates that the mean vocabulary *MA* underestimates the full scale *MA* for the foreign born children. Parenthetically speaking, we are not implying that the Binet *MA* is an accurate representation of the intelligence of these children. Frankly, we do not know.

An observation is sometimes made, whenever the vocabulary *MA* is higher than the Binet *MA*, it is merely a reflection of the fact that the *IQ* of the group is high (17). In order to test if such a relationship actually existed, we compared the Binet *MA*s of native and foreign born children to the vocabulary *MA*s by *IQ* levels. We could not compare vocabulary scores of less than five, since no *MA* equivalents are available (5).

Table 4 presents a comparison of the Binet *MA*s of the native born and foreign born children compared to the vocabulary *MA*s by the *IQ* level. While it is to be noted that the number of children at the lower *IQ* level is rather small, yet, by inspection of the table, it becomes quite evident that the vocabulary differences were approximately the same at every *IQ* level. We can thus state that, on the basis of our findings, a higher vocabulary *MA* than a Binet *MA* is not a function of the *IQ* level. It may possibly be an artifact of the socioeconomic and cultural level of the group. Children with varying *IQ*s who come from the same socio-economic background would, according to this hypothesis, tend to have approximately the same deviation between vocabulary and Binet *MA*. This is similar to the findings of Baker, *et al.* (2) who were working with a superior group of children (*IQ* 120). They indicated that children with accelerative changes in mental growth did not differ from those with declining rates of mental growth, in any area of ability as measured by the Revised Stanford-Binet Scale, Form *L*.

TABLE 4

THE BINET *MA*S OF NATIVE AND FOREIGN BORN CHILDREN COMPARED TO THEIR VOCABULARY *MA*S BY *IQ* LEVELS

	Binet *MA* higher				Binet vocabulary Higher				Binet vocabulary Less than 5 words				No difference			
	Native		Foreign		Native		Foreign		Native		Foreign		Native		Foreign	
IQ	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Below 90	1	0.16	2	1.47	5	0.78	3	2.21	9	1.41	12	8.82				
90- 99	11	1.72	4	2.94	23	3.59	3	2.21	19	2.97	22	16.18				
100-109	33	5.16	16	11.76	68	10.63	12	8.82	18	2.81	23	16.91	1	0.16		
110-119	50	7.81	9	6.62	100	15.63	6	4.41	26	4.06	10	7.35	4	0.63		
120-129	30	4.69	3	2.21	80	12.50	3	2.21	11	1.72			6	0.94		
130-139	27	4.22	0	—	48	7.50	3	2.21	4	0.63			1	0.16	1	0.74
140 & over	22	3.44	2	1.47	37	5.78	1	0.74	1	0.16			5	0.78	1	0.74
Total	174	27.20	36	26.47	361	56.41	31	22.81	88	13.76	67	49.26	17	2.67	2	1.48

Further examination of Table 4 showed that 86.28 per cent of the native born children had a vocabulary of five or more words, whereas only 50.76 per cent of the foreign born children had a vocabulary of five or more words. Even when we considered the fact that the Binet *MA* of the foreign born children was lower than that of the native born group, the vocabulary *MA* and Binet *MA* differences could not be fully accounted for by differences in *MA* alone. Obviously, other factors made the Binet vocabulary much more difficult than the Binet *MA* for the foreign born children.

Table 5 presents the differences between the Binet *MA*s of the native born and foreign born children as compared to their vocabulary *MA*s. It is to be recalled that this is a comparison of 86.28 per cent of native and 50.76 per cent of foreign born children's vocabularies.

The native born children had 31.51 per cent of the Binet *MA*s higher than the vocabulary *MA*s and 65.40 per cent vocabulary *MA*s higher than the Binet *MA*s. Furthermore, what was most striking was that, whereas 50.91 per cent of the vocabulary *MA*s were seven or more months higher than the Binet *MA*s, only 18.29 per cent of the Binet *MA*s were seven or more months higher than the vocabulary *MA*s.

In 52.18 per cent of the cases, the foreign born children had Binet *IQ*s higher than the vocabulary *MA*s and in only 44.93 per cent they had vocabulary *MA*s higher than the Binet *MA*s.

F. DISPLACEMENT IN THE ORDER OF DIFFICULTY OF THE BINET VOCABULARY

We tabulated the order of difficulty of the Binet vocabulary by ages and nativity. Since the age group eight to nine-eleven had very few cases, we felt that a separate tabulation for this group was not warranted and results for this group were included in the tabulation for the entire population.

Table 6 presents the order of difficulty of the Binet vocabulary for the native born children. Note displacement in rank. The word "juggler" changed from Rank 15 to 9 and the word "scorch" from Rank 9 to 15. The word "haste" changed from Rank 11 to 17.

Table 7 presents the order of difficulty of the Binet vocabulary for the foreign born children. Note displacement in rank of certain words. In reviewing Tables 6 and 7, we note that the relative difficulty of the words appears to be the same for both groups. Differences, when they do appear, are minor and are due to the greater unreliability of the ranks for the foreign born group as very few foreign born children knew the more difficult words.

As there has been a change in the difficulty of the words, it is advisable to

TABLE 5

THE DIFFERENCES BETWEEN THE BINET *MA*s OF THE NATIVE BORN AND FOREIGN BORN CHILDREN AS COMPARED TO THEIR VOCABULARY *MA*s

	Binet *MA* higher				Vocabulary *MA* higher				No difference			
	Native		Foreign		Native		Foreign		Native		Foreign	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
No difference									17	3.08	2	2.90
1 to 6 mo.	73	13.22	12	17.39	80	14.49	8	11.59				
7 to 12 mo.	51	9.24	12	17.39	93	16.85	11	15.94				
13 to 18 mo.	32	5.80	7	10.15	83	15.04	4	5.80				
19 to 24 mo.	11	1.99	3	4.35	40	7.25	5	7.25				
25 to 30 mo.	3	0.54	0	0.00	20	3.62	1	1.45				
31 to 36 mo.	3	0.54	1	1.45	33	5.98	1	1.45				
37 to 42 mo.	1	0.18	0	0.00	11	1.99	0	0.00				
43 to 48 mo.	0	0.00	1	1.45	1	0.18	1	1.45				
Total	174	31.51	36	52.18	361	65.40	31	44.93	17	3.08	2	2.90

TABLE 6

ORDER OF DIFFICULTY OF THE FIRST EIGHTEEN WORDS OF THE REVISED STANFORD-BINET VOCABULARY FOR NATIVE BORN CHILDREN AGED FOUR YEARS TO NINE YEARS ELEVEN MONTHS

		CA 4 to 5-11 *N* = 446			*CA* 6 to 7-11 *N* = 155			*CA* 4 to 9-11 *N* = 640		
Old rank	Word	New rank	*N*	%	New rank	*N*	%	New rank	*N*	%
1	Orange	1	443	99.33	1	155	100	1	637	99.53
2	Envelope	3	404	90.58	3	149	96.13	3	588	91.88
3	Straw	2	432	96.86	2	148	95.48	2	615	96.09
4	Puddle	4	357	80.04	5	138	89.03	4	530	82.81
5	Tap	6	331	74.22	4	137	88.39	6	501	78.28
6	Gown	5	351	78.70	6	128	82.58	5	510	79.69
7	Eyelash	8	208	46.64	8	82	52.90	8	316	49.38
8	Roar	7	217	48.65	7	102	65.80	7	348	54.38
9	Scorch	14	9	2.02	15	6	3.87	15	23	3.59
10	Muzzle	11	20	4.48	12	19	12.26	11	56	8.75
11	Haste	18	2	0.45	17	4	2.58	17	14	0.22
12	Lecture	13	10	2.24	14	11	7.10	14	40	6.25
13	Mars	10	83	18.61	9	70	45.16	10	182	28.44
14	Skill	15	7	1.57	13	16	10.32	13	41	6.41
15	Juggler	9	100	22.42	10	69	44.52	9	198	30.94
16	Brunette	12	13	2.91	11	23	14.84	12	50	7.81
17	Peculiarity	17	3	0.67	17	4	2.58	18	12	0.19
18	Priceless	16	4	0.90	17	4	2.58	16	17	0.27

TABLE 7

ORDER OF DIFFICULTY OF THE FIRST EIGHTEEN WORDS OF THE REVISED STANFORD-BINET VOCABULARY FOR FOREIGN BORN CHILDREN AGED FOUR YEARS TO NINE YEARS ELEVEN MONTHS

		CA 4 to 5-11 *N* = 72			*CA* 6 to 7-11 *N* = 57			*CA* 4 to 9-11 *N* = 136		
Old rank	Word	New rank	*N*	%	New rank	*N*	%	New rank	*N*	%
1	Orange	1	68	94.44	1	57	100	1	132	97.06
2	Envelope	3	50	69.44	3	45	78.95	3	102	75.00
3	Straw	2	62	86.11	2	50	87.72	2	118	86.76
4	Puddle	6	30	41.67	5	31	54.39	5	66	48.53
5	Tap	4	38	52.78	4	36	63.16	4	80	58.82
6	Gown	5	33	45.83	6	22	38.60	6	59	43.38
7	Eyelash	8	11	15.28	8	13	22.81	8	30	22.06
8	Roar	7	18	25.00	7	20	35.09	7	43	31.62
9	Scorch	11	1	1.39	13.5	1	1.75	15	2	1.47
10	Muzzle	—	0	0	10	5	8.77	11	7	5.15
11	Haste	—	0	0	—	0	0	15	2	1.47
12	Lecture	—	0	0	13.5	1	1.75	15	2	1.47
13	Mars	10	4	5.56	11	4	7.02	10	13	9.56
14	Skill	11	1	1.39	13.5	1	1.75	13	4	2.94
15	Juggler	9	6	8.33	9	7	12.28	9	17	12.50
16	Brunette	11	1	1.39	13.5	1	1.75	12	6	4.41
17	Peculiarity	—	0	0	—	0	0	18	0	0
18	Priceless	—	0	0	—	0	0	17	1	0.74

modify Terman's instructions (18) and, instead of discontinuing the vocabulary test after six consecutive failures, to discontinue it only after 10 consecutive words have been failed.

G. Summary and Conclusions

A study of the revised Stanford-Binet, Form *L* records of 640 native and 136 foreign born children, aged 4 to 9-11, indicated that there is a statistically significant difference between the vocabulary *MA* for the same Binet *MA* in favor of native born children. The mean vocabulary *MA* overestimates the full scale *MA* for native born children and underestimates the full scale *MA* for foreign born children. There has been a change in the order of difficulty of the words. However, the relative difficulty of the words remains about the same for both groups.

It is suggested that the examiner, in administering the vocabulary test, should continue until at least 10 consecutive words have been failed. While the Binet vocabulary is still a valid index of intelligence on the kindergarten and primary school level, as indicated by correlation with the entire scale, its use alone as a short verbal scale of intelligence, is not advisable.

References

1. Arnold, F. C., & Wagner, W. K. A comparison of Wechsler Intelligence Scale and Stanford-Binet scores for eight- and nine-year-olds. *J. Exp. Educ.*, 1955, **24**, 91-94.
2. Baker, C. T., Sontag, L. W., & Nelson, V. L. Specific ability in *IQ* change. *J. Consult. Psychol.*, 1955, **19**, 307-310.
3. Barratt, E. S., & Baumgarten, D. L. The relationship of the *WISC* and Stanford-Binet to school achievement. *J. Consult. Psychol.*, 1957, **21**, 144.
4. Cohen, B. D., & Collier, M. J. A note on *WISC* and other tests of children six to eight years old. *J. Consult. Psychol.*, 1952, **16**, 226-227.
5. Cureton, E. E. Mental age equivalents for the Revised Stanford-Binet vocabulary test. *J. Consult. Psychol.*, 1954, **18**, 381-383.
6. Elwood, M. I. A preliminary note on the vocabulary test in the Revised Stanford-Binet Scale, Form *L*. *J. Educ. Psychol.*, 1939, **30**, 632-634.
7. Feifel, H., & Lorge, I. Qualitative differences in the vocabulary responses of children. *J. Educ. Psychol.*, 1950, **41**, 1-18.
8. Frandsen, A. N., & Higginson, J. B. The Stanford-Binet and the Wechsler Intelligence Scale for Children. *J. Consult. Psychol.*, 1951, **15**, 236-238.
9. Garrett, H. E. Statistics in psychology and education, 4th Edition. New York: Longmans, Green, 1953.
10. Krugman, J. I., Justman, J., Wrightstone, J. W., & Krugman, M. Pupil functioning on the Stanford-Binet and the Wechsler Intelligence Scale for Children. *J. Consult. Psychol.*, 1951, **15**, 475-483.
11. Levinson, B. M. A comparative study of the intelligence of Jewish preschool boys and girls of orthodox parentage. *J. Genet. Psychol.*, 1957, **90**, 17-22.
12. McNemar, Q. The Revision of the Stanford-Binet Scale. New York: Houghton-Mifflin, 1942.

13. Pastovic, J., & Guthrie, G. M. Some evidence on the validity of the *WISC*. *J. Consult. Psychol.*, 1951, **15**, 385-386.

14. Pintner, R., Dragositz, A., & Kushner, R. Supplementary guide for the Revised Stanford-Binet Scale (Form *L*). *Appl. Psychol. Monog.*, 1944, No. 3.

15. Rapaport, D., Gill, M., & Schafer, R. Diagnostic psychological testing, Vol. I. Chicago: Yearbook Publishers, 1945.

16. Spache, G. Methods of predicting results of full scale Stanford-Binet. *Amer. J. Orthopsychiat.*, 1944, **14**, 480-482.

17. ———. The vocabulary tests of the revised Stanford-Binet as independent measure of intelligence. *J. Educ. Res.*, 1942-43, **36**, 512-16.

18. Terman, L. M., & Merrill, M. A. Measuring intelligence. New York: Houghton-Mifflin, 1947.

19. Triggs, F. O., & Cartee, J. K. Preschool pupil performance on the Stanford-Binet and the Wechsler Intelligence Scale for Children. *J. Clin. Psychol.*, 1953, **9**, 27-29.

20. Tuddenham, R. D. Soldier intelligence in World Wars I and II. *Amer. Psychol.*, 1948, **3**, 54-56.

21. U. S. Department of Commerce, Office of Business Economics. *National income, 1954 Edition*. Washington: U. S. Govenment Printing Office, 1954.

22. U. S. Office of Education. *Statistical summary of education*. Washington: U. S. Government Printing Office, 1953.

23. Wechsler, D. Wechsler Intelligence Scale for Children: Manual. New York: Psychological Corporation, 1949.

24. Weider, A., Noller, P. A., & Schramm, A. The Wechsler Intelligence Scale for Children and the Revised Stanford-Binet. *J. Consult. Psychol.*, 1951, **15**, 330-333.

Psychological Center
Graduate School of Education
Yeshiva University
110 West 57th Street
New York 19, New York

The Journal of Genetic Psychology, 1958, **93**, 249-276.

THE ECOLOGICAL VERACITY OF THE SELF-REPORT*

Human Development Institute, University of California, and Department of Psychology, University of Colorado

SAMUEL R. PINNEAU AND ALEXANDER MILTON[1]

A. INTRODUCTION

The instructions for most personality inventories indicate to the subject that his responses are to be regarded as assertions on his part regarding the presence or absence of the attributes to which the items refer. However, since the conventional definition of validity deals not with the truthfulness of these "item-responses" (assertions), but with the extent to which they are related to the criteria for the characteristic or trait being measured, it is usually not considered necessary that they correspond either to reality or to the subject's perceptions of reality in order for the test to be valid. In fact it may be argued that the diagnostic value of some items arises from the tendency of a particular group of subjects to bias their responses in a direction opposite both to reality and to their perception of reality.

Some writers have advocated an even more extreme point of view: that we should consider item-responses as *just* a class of verbal behavior. For example, Meehl denies that we should accept the subject's claims as a surrogate for a behavior sample and he substitutes the assertion that ". . . a 'self-rating' [item-response] constitutes an intrinsically interesting and significant bit of verbal behavior, the non-test correlates of which must be discovered by empirical means" (12, p. 297). The obvious implication of this position is that while a high score on a scale may be a quite valid indicator of relative position on a particular variable, the item-responses cannot, without additional evidence, be considered to reflect directly the individual's behaviors, feelings, emotions, interests, beliefs, etc.

Meehl gives a number of examples in which he considers that it is unlikely

---

*Accepted for publication by Harold E. Jones of the Editorial Board, and received in the Editorial Office on October 11, 1957.

[1] The initial results of this study were included in a paper presented at the 34th meeting of the Western Psychological Association, Long Beach (18). The authors wish to express their appreciation to William F. Vibrans of Eugene, Oregon, for his assistance in the collection of data, to Ruth A. Pinneau and Marilou Milton for their assistance in the data analysis, and to Professor Harold E. Jones for his suggestions and criticisms in the writing of the manuscript.

[2] At the time of the final analysis of the data, the first author, located at the Institute of Child Welfare, University of California, was a Public Health Service Research Fellow of the National Institute of Mental Health, and the second author was located at Stanford University.

that the subject's responses agree with reality; e.g., he points out the psychopath's tendency to respond "True" to such items as "My parents and family find more fault with me than they should." He suggests that the discrepancy is to be accounted for in terms of conscious and unconscious "fudging," in terms of differential interpretation of the items, and/or in terms of the subject's inability to evaluate reality objectively.

Although it may be true that the examples cited by Meehl represent distortions of objective reality, the question may be raised as to whether Meehl has appropriately evaluated his examples since the instructions on such tests as the *MMPI* seem to imply that the individual should answer the questions in terms of his *perception* of the world. It seems quite plausible to the present investigators that the correspondence between the subject's item-responses and his *perception* of reality may be high in the above example and in others provided by Meehl. If such is the case, it may be that the usefulness of many items lies in their power to determine how the subject sees the world and not in the correspondence of the subject's item-responses to reality. If this is so, it would seem that lack of agreement between the subject's claims and objective reality would not be a sufficient reason for treating the subject's item-responses only as a class of verbal behavior and thereby discarding the possible information gained about the subject's perceptual world.

Lack of agreement between the subject's responses and his perception of reality could perhaps be accounted for in essentially the same terms as Meehl uses to account for the discrepancy between his responses and objective reality.

The relationship of the self-report to the "true state of affairs" has also been considered by Goodenough who has provided a statement which perhaps represents the view held by a majority of contemporary psychologists. She states: "Paper-and-pencil tests . . . do not always yield dependable information about the actual behavior of the respondents in the situations described. Strictly speaking, all that they provide is a series of claims which may or may not conform to the behavior under consideration" (3, p. 383).

There would be considerable advantage in being able to assume in interviews, questionnaires, and personality inventories that the subject's responses correspond (*a*) to objective reality or (*b*) to his perception of reality. To the extent to which the first assumption is true, an experimenter would be able to obtain actual or "reality" correlates of the trait and, to the extent to which the second assumption is true, the experimenter would obtain perceptual correlates. In view of the importance of information about the subject's objective world and his perception of it, it would seem desirable to determine the extent to which these assumptions are tenable. Until their validity has been investigated it would seem undesirable to conclude that test responses are *just* verbal behavior.

B. The Concept of Veracity

As was pointed out earlier, the conventional concept of validity does not encompass the correspondence of the subject's report either to objective reality or to his perception of reality. In order to designate this correspondence, the term *veracity* will be used.

To the extent that a subject's response corresponds to his perception of the world, it may be said to have *personal veracity* or truthfulness, and to the extent to which the response corresponds to objective reality, it may be said to have *ecological veracity* or truthfulness. An empirical estimate of the correspondence can be obtained by comparing the self-report with observational records: Ecological veracity may be estimated by comparing the subject's report with an external observer's report of the "true state of affairs"; and personal veracity may be estimated by comparing the subject's report concerning his perceptual world with the observer's descriptions of it. (In the latter case, the descriptions would, of course, represent judgments based on behavioral observations.)

Perhaps the distinction between the two kinds of veracity will be made clearer by considering an extreme example, the case of a paranoiac who answers "True" to the item "I am being plotted against." A comparison of the item-response with the observers' reports of the "true state of affairs" would indicate that the item-response has no ecological veracity, assuming that the observers agreed that this was not the case. On the other hand, assuming the accuracy of the observers' reports, the item would have personal veracity if they agreed that the paranoiac acted as though he perceived that he was being plotted against.

In instances in which the estimated veracity falls below the maximum possible value, it would be necessary to assume that the subjects' reports lacked veracity, unless it could be shown that it was the observers' reports which were inaccurate.

C. Problem and Method

The present study is concerned with determining the correspondence between the subject's report of simple everyday behaviors, feelings, emotions, attitudes, and interests and the report of an observer who has had ample opportunity to observe the subject's behavior and expression of feelings, emotions, and attitudes in representative life situations; i.e., the present study represents an attempt to estimate the *ecological veracity* of the subject's self-report. (Since the concept of personal veracity will not be considered again in the remainder of this article, the term veracity will be used as a synonym for ecological veracity.) Three aspects of veracity will be considered: Methods of estimating veracity, the veracity of different kinds of items, and variables related to subject-observer disagreements.

The correspondence between the reports of subjects and observers on psychological tests can be measured in several different ways, two of which will be employed in the present study: (*a*) The correspondence can be expressed in terms of percentage of agreement, i.e., in terms of the percentage of overlapping responses. One of the obvious limitations of this measure is that even if the percentage of agreements were high for a given sample, no relationship would necessarily be present between the directional answering of subjects and observers. If no correlation were present, the agreement could be accounted for as adequately by response sets common to subjects and observers as by veracity on the part of the subjects. While this possibility must be recognized, it would seem unlikely in instances in which observers are asked to report only on reactions on which they have information.

Another relevant limitation of this measure is that in the absence of two self-reports and two observational records there would be no basis for estimating the maximum amount of agreement ("maximum amount of veracity") that the reliability of the records would permit.

(*b*) The correspondence of the reports can be expressed by the correlation between the subjects' and observers' *test scores*. These correlations, termed *veracity coefficients*, can be compared with the observers' and subjects' reliabilities to determine how score agreement between subject and observer compares with the consistency of the reports or they can be compared with the maximum possible relationships estimated from these reliabilities. While the veracity coefficients do not directly reflect the proportion of overlapping responses, in conjunction with other information they can be used to show the extent to which any tendency to get comparable scores is to be accounted for in terms of the subjects and observers answering the same items in the same direction, and, under certain conditions, they indicate whether the percentage of agreement approximates the maximum possible value permitted by the reliability of the reports.

The veracity of the self-report may depend on the kind of item used. To determine if this is the case, the items used in the present study were classified according to three different schemes and the veracity of each was determined:

1. *Psychological variables.* The items were grouped according to the

psychological variables to which they are purportedly related.

2. *Response form.* All items were classified according to whether they referred to behaviors, to interests, or to beliefs, feelings, emotions, etc. While it would have been desirable to determine separately the veracity for beliefs, feelings, and other assumed latent structures, there was insufficient agreement among judges to warrant the breakdown of the items into such finer groupings.

3. *Response frequencies.* Reactions may be grouped not only on the basis of their form or according to the variable to which they are related, but they may also be classified as to their frequency of occurrence. Hence, scales for extreme responses were established on the basis of the frequency with which the items were answered as "True" in a normal population.

A number of factors may result in disagreement between subject and observer. Adequate checks on these factors would include checks on the reliability of the reports of subjects and observers, on the subject's ability to evaluate reality objectively, on his tendency to fake on certain of the items, and on his or the observer's tendency to interpret some of the items in an unusual way, etc. With the present data it is not possible to investigate all of the variables which might account for disagreement between subject and observer; it is possible, however, to check a number of hypotheses regarding the causes of disagreement:

1. It might be expected that unreliability of the reports of subjects and observers would account for a considerable proportion of the disagreements.

2. It might be expected that an important factor would be variations in the length of time covered by the observation.

3. It might be expected that the observer who answered all of the items would be less accurate than the observer who was more selective in answering the items; i.e., one might hypothesize that the more selective the observer, the greater would be his agreement with the subject.

4. In the light of Cronbach's article (2) on processes affecting scores, it might be expected that disagreement would be related to relative standing on some of the scales or that the observers might tend to bias their answers towards the mean, resulting in greater disagreement in the case of the more extreme subjects.

5. Cronbach (2) also suggests that there would be higher disagreement on the more deviant items, i.e., on those items which are answered in a given direction by a smaller percentage of a normal population; hence the larger the number of deviant responses shown by a subject, the greater the disagreement.

D. PROCEDURE

1. *Subjects and Observers*

The subjects were 45 married male students at the University of Oregon, the majority of whom lived in one of the University's housing projects. Most of the subjects participated in the study in response to a door to door request. The observers, who had ample opportunity to observe the subjects in a large number of representative life situations, were the subjects' wives.

In all but three cases either the subject or his wife participated in the study at the University on one of two evenings. In the case of the couples in which one member of the pair had to stay at home, a test was delivered to the home at the same time the spouse was at the University and it was picked up before he returned. In three instances in which the couples were together during a portion of the test, precautions were taken to insure that they did not discuss the experiment until after the protocols were turned in.

At the University, the subjects were tested in one room and the observers in another, care being taken that neither group knew what the other group was doing.

2. *Materials and Test Instructions*

The test used with the subjects consisted of three scales from the Minnesota Multiphasic Personality Inventory (*MMPI*) (8) and three comparable scales from the Multi-Dimensional Behavioral Inventory, a test in the experimental stage of development (17). The *MMPI* scales were the *M-F* scale, the scale of social introversion (*Si*), and Gough's social status scale (*SS*) (a total of 140 different items). The Behavioral Inventory scales were the maleness-femaleness scale (*M-F*) (18), the need for social contact scale (*SC*) ("social introversion") (14), and the social status scale (*SS*) (15) (a total of 334 different items). In contrast to the latter inventory which consists only of items of behavior, the *MMPI* consists of items of behavior, interests, feelings, emotions, attitudes, and beliefs. Three sample items from each of the scales of the Behavioral Inventory are given in Table 1.

The subjects were told: "The purpose of asking you to take the test is to determine how many times each item is answered in a given way by a group of married males and how such answers correspond to your wives' responses." They were requested to answer every item as either "True" or "False" and were assured that their test responses would be kept confidential. The description of the nature of the study was purposively ambiguous in order not to be obviously deceptive while at the same time disguising the purpose of the study. Subsequent to the test a check was made to determine if the subjects realized that the accuracy of their reports was being investigated. These interviews indicated that they were not aware that their wives were reporting on their behavior but thought that the wives were also giving self-reports.

TABLE 1
SELECTED ITEMS FROM THE THREE MALE SCALES OF THE BEHAVIORAL INVENTORY

Social Contact Scale

Subjects low on social contact tend to answer these items in the indicated direction:

1. When eating peas, corn, etc., I sometimes use bread as a pusher. (T)
2. I sleep in the nude. (T)
3. In a restaurant I usually sit at the counter. (F).

Social Status Scale

Subjects of low social status tend to answer these items in the indicated direction:

1. At a cafe or restaurant I sometimes wipe my silverware with my napkin. (T)
2. I usually carry paper book matches. (T)
3. I frequently take aspirins or anacins several times a week. (T)

Maleness-Femaleness Scale

Male subjects answer these items in the indicated direction more often than females:

1. I frequently drive with only one hand on the wheel. (T)
2. I often tell jokes during a conversation. (T)
3. I always use cream and sugar in my coffee. (T)

The test for the observers consisted of the same items but they were couched in the third person. They were told that: "The purpose of asking you to take the test is to determine how much you are aware of your husband's feelings and habits." They were requested to answer every item as either "True," or "False," or "Do not know." The instructions specified that they should include in the latter category the behaviors which they had not observed and the feelings, attitudes, etc., of which they did not know.

E. Results[3]

The record of two of the couples had to be excluded from the results, one because the wife failed to comply with the instructions in answering all of the questions in one of the three specified ways, and the other because the wife, after responding to approximately three-quarters of the items of behavior, refused to answer the remainder except by "Do not know." Hence, the reported results are based on 43 couples.

---

[3] Except for the most basic data, the authors have omitted tabular material on the individual scales, indicating by median values the trend of the results. Tables presenting the results of all the indicated analyses can be obtained either from the authors or from the American Documentation Institute. Order Document No. 5457 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Project, Library of Congress.

1. *Agreement between Subject (Husband) and Observer (Wife)*

On the average the observers answered 96 per cent of the items from the Behavioral Inventory as either "True" or "False." On these items the subjects agreed with the observers 78 per cent (median) of the time. The observers were willing to answer a significantly smaller percentage of the *MMPI* items as either "True" or "False" and of those thus answered they agreed significantly less often with the responses of their husbands (median= 73 per cent). Table 2 presents means, medians, and quartiles for these measures. Because of the skewed distributions, the sign test was used to

TABLE 2
PERCENTAGES OF ITEMS ANSWERED AND PERCENTAGES OF AGREEMENT

A. All scales

	Per cent answered BI	Per cent answered MMPI	Per cent agreed BI	Per cent agreed MMPI
Mean	95.5	91.0	77.5	74.0
Median	96	93	78	73
Q_1	94	86	75	69
Q_3	98	96	80	76
Chi square (sign test of difference between tests.)		28.44***		16.10***

B. Individual scales

Behavioral Inventory

	Per cent answered: Chi squares (sign test of difference between scales.)		Mean	Median	Q_1	Q_3		Per cent agreed: Chi squares (sign test of difference between scales.)		Mean	Median	Q_1	Q_3
MF			94.1	95	92	98	MF			77.5	78	74	82
	10.94***							0.40					
SS		11.92***	95.2	97	94	98	SS		2.50	76.6	76	74	80
	7.80**							5.76*					
SC			96.0	98	95	99	SC			79.1	79	76	83

MMPI

	Per cent answered: Chi squares (sign test of difference between scales.)		Mean	Median	Q_1	Q_3		Per cent agreed: Chi squares (sign test of difference between scales.)		Mean	Median	Q_1	Q_3
MF			87.5	90	82	96	MF			74.3	74	71	78
	4.24*							8.80**					
SS		21.78***	90.1	94	88	97	SS		1.20	80.2	80	75	87
	5.44*							18.66***					
SI			92.0	96	87	99	SI			72.1	73	67	77

*Values significant at .05 level.
**Values significant at .01 level.
***Values significant at .001 level.

estimate the significance of the differences between scales. Use of the "Don't know category" by the observers varied from 0 to 22 per cent in the case of the Behavioral Inventory, and from 0 to 50 per cent on the *MMPI*.

It was earlier suggested that the veracity of the self-report may vary from one area to another. In order to determine if this is the case the percentage of agreement was determined for the item correlates of the three variables (social status, masculinity-femininity, and social contact or social introversion). The results of this analysis, presented in Table 2, show that significant differential agreement was obtained between the two inventories and in some instances between the individual scales.

2. *The Relationship between Scores on the Test Scales*

To obtain an estimate of the degree of relationship between the subjects' scores and those of the observers, it was necessary to score each of the husbands' protocols on the six scales for the same items which their wives had answered. Thus, the scores are based on "part scales." The wives' protocols were also scored.

It might be argued that the use of a varying number of items per subject-observer pair would inflate both the correlations between husbands' and wives' scores and the reliability of these scores. If so, scale elevation would be related to number of items answered. For both inventories, this was the case; the median correlations were .24 for the Behavioral Inventory and .65 for the *MMPI*. Because of this relationship it was necessary to correct the individual scores so that they would be based upon scales of uniform length. For each scale this correction involved adding to the subjects' and observers' scores that proportion of the unanswered items of the scale which the "part scale" scores were of the number of items both answered. Thus proportional answering was assumed for the subjects; in the case of the observers it was assumed that they would have answered a proportional number of the unanswered items in the same direction if they had had the same opportunity to observe these reactions.

After the scores were corrected to full-scale length, the median correlations between score elevation and number of items answered were reduced to —.16 for the Behavioral Inventory and .23 for the *MMPI* and thus the correction appears warranted. Also supporting the validity of the procedure was the increase in the correlations between the subject's full scale scores and his scores on the "part-scales." The median correlation increased from .85 to .98 for the Behavioral Inventory and from .81 to .95 for the *MMPI*. Thus the subjects' corrected "part-scale" scores very closely approximate their relative standing on the full scales. The relationships between the subjects' full scale scores and observers' "part-scale" scores suggest a similar trend; the median value increased from .40 to .53 for the Behavioral Inventory, and from .42 to .46 for the *MMPI*.

Correlations were computed between the husbands' and wives' scores thus corrected (cf. Table 3). These veracity coefficients could be compared with the reliabilities for the scales reported in a number of studies (8). Such comparisons would be misleading, however, since the subjects in this study represent a relatively homogeneous group; therefore, the writers estimated the reliabilities of the trait measures by the Spearman-Brown formula. The reliability estimates are presented in Table 3. [Before computing the correlations between the halves of the scale (alternate items) the scores were corrected to scales of uniform (half) length.] Assuming that these coefficients are adequate estimates of the reliability of the reports, they were used to determine the *maximum possible* correlation between the subjects' and observers' scores (cf. Table 3) (11).

In evaluating the difference between the veracity coefficient and the maximum possible relationship the writers used the standard error of the veracity coefficient. This procedure is subject to the criticism that the maximum possible relationship is being treated as a hypothetical value when it is, in fact, an obtained value and not independent of the same factors which influence the veracity coefficient. In this case, however, evidence will be presented which indicates that the errors are not positively correlated. Therefore if lack of independence has any influence, it is to work against the authors' hypothesis.

Table 3 indicates that in one instance the "maximum possible" relationship is significantly greater than the veracity coefficient and that in one instance the obtained veracity coefficient is actually higher than the square root of the product of the reliability coefficients. There may be question as to how much importance should be attached to these differences: While estimates of reliability by the split-half procedure yield results consistent on the average with those obtained by comparable-form procedures, the dispersion of split-half reliabilities is greater.[4] Hence, the dispersion of "maximum possible" relationships based on these would be greater than when based on comparable-form reliabilities. This may partially explain why two apparently

---

[4] The estimate of reliability obtained by the Spearman-Brown formula is too low when the halves of the test are not comparable. It is too high when correlated variable errors are present. The assumption that the reliability coefficients (obtained by this formula in the present study) are adequate estimates of reliability is in essence an assumption that on the *average* these factors counterbalance each other.

TABLE 3

COMPARISONS OF SUBJECTS' AND OBSERVERS' CORRECTED SCORES

	Veracity coeff.	Maximum Veracity coeff.	z†	Reliability coefficients Subject	Reliability coefficients Observer	Means Subject	Means Observer	*CR*	SD's Subject	SD's Observer	t (Variances)
	Behavioral Inventory										
MF	.38	.51	1.03	.54	.48	95.4	88.6	4.79**	7.98	8.59	.52
SS‡	.58	.20	2.91**	.26	.16	58.5	58.6	.18	6.56	6.41	.17
SC	.59	.64	.51	.80	.51	40.1	38.1	2.07*	7.32	6.17	1.09
	MMPI										
MF	.52	.65	1.25	.55	.76	28.4	27.0	1.76	4.54	5.41	1.33
SS	.52	.31	1.62	.39	.25	24.8	24.5	.68	3.13	2.85	.70
Si	.44	.80	3.96**	.76	.85	21.0	22.3	.95	6.67	8.98	2.16*

*Differences significant at .05 level.

**Differences significant at .01 level.

†*z* is the critical ratio comparison of the difference between the veracity coefficient and the maximum possible veracity.

‡The reliability coefficients for the social status scale of the Behavioral Inventory are a great deal lower than those previously obtained with the scale. This reduction appears to be explained in part by the restricted "range of talent"; the *SD* for the original group was approximately 2⅓ to 2½ times larger than those obtained with the present subjects and observers. It may also be that the scale is an inadequate measure of social status for this married sample since it was derived on a group of younger single males.

significant differences in opposite directions were obtained. On the basis of this analysis it appears reasonable to conclude that for all practical purposes there is as high a relationship between the subjects' and observers' scores as the reliabilities of the measures permit.

It might be argued that the preceding conclusion assumed that the variable errors of the subjects and observers are not positively correlated, i.e., that there are no common response sets, and that this assumption is questionable since the subjects and observers are husbands and wives. However, if correlated errors are present, they must necessarily affect the estimates of the maximum possible relationships since these were based on split-half reliabilities which are even more open to correlated variable errors, involving as they do subject-subject and observer-observer correlations. Under such conditions, the presence of correlated variable errors would result in the maximum possible coefficients exceeding the obtained relationships. Since the obtained coefficients approximate the maximum possible values it would appear that correlated variable errors have not contributed to any great extent to the magnitude of the reliability and veracity coefficients.

The scores of the subjects and observers were obtained on the same items while the scores used in computing the reliability coefficients were obtained on different items. Hence, it might be argued that response sets common to subjects and observers would contribute as much to the veracity coefficients as correlated variable errors in the halves of the tests would contribute to the reliability coefficients. If response sets common to observer and subject contribute to any great extent to the magnitude of the veracity coefficients, one would expect the observers' scale scores on items on which they disagreed with the subjects to be positively related to the subjects' scores on the total scales. For the Behavioral Inventory these correlations ranged between —.34 and —.52, and for the *MMPI* between .14 and —.51. Thus, in general, the observers' responses on the disagree items are *not* in the trait direction characteristic of the subject.

In order to determine the consistency with which the observers tended to answer the disagree items in a given scale direction, the correlations between the disagree scores for the two halves of each scale were computed. The values ranged from .00 to .69. (The median value was .36 for the Behavioral Inventory and .17 for the *MMPI*.) These relationships indicate the operation of correlated variables in the observers' responses to the disagree items in the halves of the scales. (The median correlation between the subjects' and observers' scores on the disagree items was —.13; the range was .18 to —.27.)

The effect of the negative relationship between the observers' disagree scores and the subjects' trait scores is a reduction in the magnitude of the veracity coefficients which would be present if the relationships were absent or positive. On the other hand the positive correlation between the observers' scores on the half scales tend to elevate the reliability coefficients and, consequently, the maximum possible relationships estimated from them. On the basis of these findings a significant difference between the veracity coefficients and the maximum relationships might be expected; however, the contribution of scores on the disagree items to total scores is so small as compared with that of common responses that the veracity and reliability coefficients are not materially affected.

3. *The Relationship between Scale Scores and Percentages of Agreement*

The preceding results indicate that in general the subjects' and observers' scores are as highly related as the reliability of the reports permit. While these findings indicate that the scores of the subjects and observers are comparable in a relative sense, the two groups could differ in the average and variability of number answered in a particular direction. The means and standard deviations of the scale scores are presented in Table 3. While for any given scale the variances did not differ significantly, the differences between the means were significant in the case of the *MF* and *SC* scores of the Behavioral Inventory: On the average the subjects reported six more male behaviors and two more low social contact behaviors, i.e., two to four more per 100 items.

If the observers had obtained scores comparable to those of the subjects by answering different items in the trait direction characteristic of the subjects, there should be a positive relationship between the observers' scale scores on the disagree items and the subjects' scale scores on the items which both had answered. The correlations were all insignificant or negative. The median values were —.37 for the Behavioral Inventory and .05 for the *MMPI*. This finding indicates that the observers obtained scores comparable to those of the subjects by answering the same items in the same direction and that the effect of the observers' responses on the disagree items was to make the scores uncomparable in an absolute sense; an effect, however, which did not significantly alter the variabilities or the means except in the case of the two scales for which the means differed by two to four responses per 100 items.

These results appear to have a significant implication for the percentage of agreement scores, namely that these percentages approximate the maximum amount of agreement permitted by the reliability of the records: Since in general the scores of the subjects and observers are as comparable as the reliabilities of the records permit, higher agreement between subject and observer would not be possible since this would result in a higher correlation between the scores, a contradiction of the finding that these values already approximate the maximum possible values.

If these percentages represent the maximum amount of agreement permitted by the reliabilities of the reports, the question might be raised as to how these values compare to those which would have been obtained between the observers of the present study and a second set of comparable observers. It would appear, for the scales reported in this study, that essentially the same values could be expected since the correlations between the scores of subjects and observers approximate the maximum possible values and since the means, variabilities, and reliabilities of the subjects' scores approximate those of the observers; hence they approximate those which we would expect to obtain from comparable observers.

F. Disagreements between Subjects and Observers

Comparisons of the veracity coefficients of the conventional scales and the "maximum possible" relationships permitted by the reliability of the reports suggest that these measures of veracity are approximately as high as is possible for the present sample. Under such circumstances, it might be expected that only small and insignificant amounts of the disagreement between subject and observer would be explained unless at the same time sources of unreliability of the scores are discovered. This conclusion assumes that the split-half reliabilities are not underestimates of the true reliabilities. To meet the contingency that they are under-estimates and, if possible, to explain disagreements between subject and observer on scales where veracity does not reach the maximum, the hypotheses regarding disagreement earlier set forth were investigated.

1. *Response Form*

The percentages of agreement were determined for all 474 items grouped according to response form, that is, according to whether they referred to behaviors, to interests, or to feelings, emotions, etc. The median percentages were 78, 72, and 75, respectively (cf. Table 4-A). The results of the comparisons of items thus grouped, presented in Table 4-B, suggest that the self-report of feelings, interests, emotions, beliefs, etc., has greater veracity and observability than has frequently been assumed while at the same time suggesting that observable behaviors can be reported with significantly greater agreement.

TABLE 4

COMPARISONS OF THE THREE CLASSES OF RESPONSE FORM

	Per cent Answered			Per cent Agreed		
	Behaviors	Interests	"Others"	Behaviors	Interests	"Others"
			A. Percentages			
Mean	94.9	86.7	90.6	77.4	73.0	74.4
Median	96	91	93	78	72	75
Q_1	93	83	86	75	68	70
Q_3	98	95	98	80	81	80
			B. Chi Squares			
				Per cent answered x^2		Per cent agreed x^2
Behaviors vs. Interests				16.9***		5.5*
Behaviors vs. "Others"				9.2**		3.0
Interests vs. "Others"				5.8*		1.6

*Values significant at .05 level.
**Values significant at .01 level.
***Values significant at .001 level.

2. *Response Frequencies*

On inventories such as the ones used in the present study, the items are answered as either "True" or "False." Two types of deviant responses may be distinguished: (*a*) The subject may indicate by his responses an *absence of usual* behaviors, beliefs, feelings, emotions, etc.; for example, in this sample a subject's answering "False" to the item, "I read the Sunday comic strip every week." (*b*) The subject may indicate by his responses the *presence of unusual* behaviors, beliefs, feelings, emotions, etc.; for example his answering "True" to the item, "I usually sleep with my head under the covers." The lack of the usual reactions may be termed *negative deviancy* and the presence of unusual ones, *positive deviancy*. That these may be two different dimensions was suggested by an unpublished study with the Behavioral Inventory in which men's scores on positive and negative deviancy scales were correlated with selected *MMPI* scales. The positive deviancy scales correlated —.17 with *Hy*, .42 with *Sc*, and .30 with *Ma*, while the negative deviancy scores correlated .34, .17 and —.12 with these scales, respectively. Using the frequency counts obtained in this earlier study, positive and negative deviancy scales for both of the inventories were established for the portions of the tests used in this study.

Each of the four deviancy scales consists of 20-22 per cent of the items used from the respective inventories. The veracity coefficients of these scales and the inter-relationships between positive and negative deviancy are given in Table 5. For both inventories the relationships between the subjects' positive and negative deviancy scores are significantly less than the possible relationship and thus provide further evidence that the distinction is meaningful.

The veracity coefficients of the *MMPI* deviancy scales (cf. Table 5) fall consistently below the maximum possible relationships, suggesting that these are the items of the *MF, SS,* and *Si* scales of the *MMPI* which have little veracity. However, as will be indicated later, an alternative interpretation may also be suggested—that these results, because of the nature of the deviancy scales, indicate less accuracy of the observers' reports.

3. *Length of Observation*

It might be expected that agreement between subject and observer would be related to length of observation as measured in this study by length of marriage. The mean length of marriage was 38 months; the range was from 8 months to over 8½ years (*SD* 23.3 months). For *number* of agreements and length of marriage the correlations were respectively .12 for the Behavioral Inventory and .13 for the *MMPI*. Similar insignificant relationships were obtained for percentage of agreement, .17 and .16 respectively. These small correlations suggest little relationship between agreement and length of observation in this sample. However, it cannot be concluded from these results that the observers' amount of information is unrelated to length of observation since they were instructed to answer only those items which they felt they had sufficient information to answer accurately. If a relationship is present between amount of information and length of observation, the number of items answered should be positively related to length of marriage. (Number of items answered, of course, would be expected to reflect not only amount of information, but also selectivity of answering. Selectivity will be considered in the next section.) For the Behavioral Inventory the correlation between length of marriage and number of items answered was .09 and for the *MMPI* .12. These results give little support to the hypotheses that in this study length of observation is related to amount of agreement or to amount of information. Since the data of this study indicate that the observers have a considerable amount of information on their spouses, it must be that much of the knowledge was obtained either prior to marriage or in the first few months afterwards.

TABLE 5

COMPARISONS OF SUBJECTS' AND OBSERVERS' SCORES ON THE DEVIANCY SCALES

	Behavioral Inventory						*MMPI*					
				Reliability coefficients						Reliability coefficients		
	Veracity coeff.	Maximum Veracity coeff.	*CR*	Subject	Observer	—Dev.	Veracity coeff.	Maximum Veracity coeff.	CR	Subject	Observer	—Dev.
+Dev.	.43	.42	.08	.64	.28	.26	.02	.51	3.43*	.55	.47	.44
—Dev.	.36	.56	1.62	.65	.48		.37	.68	2.78*	.60	.78	

*Differences significant at the .01 level.

4. *Cautiousness and Agreement*

Earlier it was suggested that the more selective (cautious) the observer was in answering the items, the greater would be his agreement with the subject, i.e., the greater his use of the "Don't know" category, the greater the agreement. This relationship was determined for each of the six conventional scales and four deviancy scales. In three instances the correlations exceed the .05 level of significance in the predicted direction. While the other seven correlations taken individually did not reach significance, it should be noted that of the total of 10 correlations, nine were in the predicted direction. The probability of such an event occurring by chance in 10 independent tests is approximately one in 100; however, since in this study the measures are not independent this finding can only be considered to suggest corroboration.

These results suggest that the more selective or cautious the observer was in answering the items, the greater his agreement with the subject. This interpretation indicates that veracity coefficients tend to represent underestimates of the "true" veracity of the self-report when uncautious or unreliable observers and/or lack of information are involved.

5. *Scale Elevation*

The hypothesis was suggested that percentage of disagreement is relatec to trait level. This hypothesis was supported in the case of the *MMPI* scales the relationships exceeding the .01 level in three of the five instances. Th question arises as to whether the relationship between disagreement and scale score is to be taken as an indication that the veracity of the self-report decreases with score, or that observation is less accurate for the subjects with higher scores. It may be that the observers' accuracy is less in the case of the introverts because their reactions are less open to observation and in the case of subjects scoring higher on *MF* because they conceal those of their less "masculine" tendencies which are tapped by such *MMPI* items as "I am strongly attracted by members of my own sex." The significant relationship between the social status scores and amount of disagreement was negative, more disagreement for the low social status subjects. This finding is less readily interpreted in similar terms unless it seems reasonable to assume that the lower social status subjects living in this middle class college environment attempt to "conceal" their lower class behaviors, feelings, emotions, etc.

Another hypothesis is that the percentages of agreement are related to deviation from the mean of the sample (2). If this were true, one would expect to find a curvilinear rather than a linear relationship between scale scores and percentage of agreements. To test this hypothesis etas were computed and the significance of departure from linearity examined in those instances in which the etas were significant. In none of the three instances in which etas were significant was the departure from linearity significant.

6. *Deviancy Scales*

The hypothesis was suggested that less agreement may be found on the items answered as either "True" or "False" by only a small percentage of the population. The results with the *MMPI* deviancy scales appear to support this hypothesis: The veracity coefficient for the corrected scores of the positive deviancy scale was .02 and for the negative deviancy scale, .37; in both instances these were less than the maximum possible relationships as estimated by the split-half reliabilities.

The question is raised as to whether these lower relationships indicate that for the deviancy scales of the *MMPI* the responses of the subjects lack veracity or whether they indicate a lack of accuracy on the part of the observer. If some of the reactions represented by the items of these scales were less open to observation either because the subjects attempted to hide them or because they were less frequently expressed, the reports of the less selective (cautious) observers would be expected to be less accurate; i.e., the number of items which they answered should be negatively related to percentage of agreement for these scales. This was the case; the correlations were —.36 and —.33 for the *MMPI* positive and negative deviancy scales, respectively.

While the veracity coefficients for both of these scales were less than the "maximum possible," only in the case of the positive deviancy scale was there an insignificant relationship between the self-report and the report of the observer. An inspection of the items of this scale suggests that it contains a considerable number of reactions that it is unlikely that male subjects would be willing to express openly in their everyday behaviors, e.g., the reactions indicated by such items as "I am strongly attracted by members of my own sex," "I would like to be a nurse," "I am embarrassed by dirty stories," and "I am bothered by people outside, on streetcars, in stores, etc., watching me," etc.

To determine if this positive deviancy scale contains a larger proportion of less observable reactions, 12 judges compared (by paired comparison) the observability of a random sample of 10 items from it with a random sample of 10 other items from the same inventory. Similar comparisons were made for the *MMPI* negative deviancy scale and for the deviancy scales of the

Behavioral Inventory. The only deviancy scale for which the items were significantly less observable than a random sample of items from the same inventory was the *MMPI* positive deviancy scale (chi square=4.69).

Knowing that positive deviancy score on the *MMPI* has a low veracity coefficient, it would be expected, if the more deviant reactions are less observable, that the higher the score, the greater the disagreement. This is the case; the correlation between score on this scale and percentage of disagreements was .65. Consistent with the finding that the items of the other deviancy scales have greater observability and veracity is the finding that for them the relationships between scores and disagreements were much smaller (median *r*, .20).

The preceding analysis suggests that there was a tendency for some of the observers to answer the items of the *MMPI* deviancy scales on the basis of insufficient information. This was especially indicated in the case of positive deviancy. Hence it may be unsafe to assume in the case of these scales that the veracity coefficients are adequate estimates of the "true" veracity of the subject's self report.

G. Discussion

A number of writers have pointed out that for empirically derived scales it is not necessary to assume a correspondence between the subjects' "claims" and the "true state of affairs" in order for the test items to be valid; however, if the subjects' assertions do correspond to reality, valuable information is being discarded if this assumption is not made. The present study was designed to determine the correspondence of the subject's self-report with his actual behaviors, interests, feelings, emotions, etc., as estimated by the correspondence between subjects' and observers' reports. The results in this study indicate that the self-report is not to be regarded as a "feeble surrogate for a behavior sample" (12, p. 297); instead they indicate that in general the self-report of these subjects was an adequate surrogate for the reports of the observers. The analysis of the data suggests that the reliability of report is the chief factor which limits this correspondence. (The correlations between the scores of the subjects and observers for three conventional scales of the *MMPI* and of the Behavioral Inventory were approximately as high as their reliabilities permitted.)

In the present study the responses of the subjects not only indicate their relative standing on the variables considered but for each subject they reveal which specific behaviors, and the expressions of which specific interests, feelings, emotions, etc., probably are and probably are not characteristic of him.

This conclusion is in marked contrast to the point of view that the subject's assertions regarding the presence or absence of the attributes are to be regarded only as a class of verbal behavior, for the results indicate that in general the item-responses of these subjects are substitutable for the observers' reports of these reactions.

1. *The Reality Correlates of Psychological Variables*

In the past, research with inventories has had as its chief function the derivation of scales for measuring psychological variables. Once derived, the scales have been considered *only* as a means for determining relative standing on these variables. Undoubtedly this approach reflects to no small extent the tendency to consider the subject's item-responses as just a class of verbal behavior. The results of the present study suggest a major expansion in the use of the inventory, namely its use as an economical means of determining the everyday behaviors and the expressions of interests, attitudes, beliefs, feelings, etc., which are the most likely "non-test" or *reality* correlates of psychological variables.

If the inventory procedure is to be used as a method for determining the trait correlates which have the highest probability of actually characterizing groups of subjects, it would be necessary to include a check on the veracity of the self-report. Such a check, it would seem, should be an integral part of the standardization procedure. In so far as possible it would seem desirable to determine in what areas the responses of subjects lack veracity even when these responses are unrelated to the criteria. This procedure would suggest the kinds of items which lead to differential interpretations and/or the areas in which the subjects are unable to evaluate reality objectively or in which they tend to "fake."

In contemporary use the inventory procedure has most often utilized a *variety* of simple everyday behaviors, interests, beliefs, feelings, emotions, etc. The present results suggest that the self-report of these reactions corresponds as closely with the report of an observer as one would expect two observers' reports to agree. In view of this it appears that the inventory approach might profitably be investigated as a means of determining the *frequency* of behaviors, and the frequency of expression of feelings, emotions, etc. If self-reports in terms of *frequency*[5] are found to have as high a veracity as has been obtained for *variety,* the inventory would seem to be a useful method of investigating the frequency with which different means are used to express different emotions, the frequency with which different means are used to satisfy different needs or motivations, etc.

---

[5] A test of the veracity of the report of casually observed frequency of behavior is found in a study by Hopper and Pinneau (9). They found that the relationship between frequency of regurgitation in the infant as reported by the mother on a questionnaire and the frequency of recorded regurgitation is almost as high as the relationship between frequency or regurgitation recorded in two adjacent days and/or weeks.

2. *Differential Veracity*

The measures of ecological veracity used here are only estimates of the true extent of correspondence between the subject's responses and reality. When these measures are low they may indicate that the subject's report has little veracity in these areas, that the observational report is inaccurate or that one or both have little reliability. The analysis of the present data indicates that where there is disagreement, part of it can be attributed to the observer's reporting on the basis of insufficient information. This finding indicates the importance of including appropriate checks on the accuracy of the observational record when the disagreement between subject and observer is to be taken as a measure of lack of veracity.

In the present study rapport appeared good and the subjects were assured that their test responses would be kept confidential. In the situation where job classification and advancement, or diagnosis and treatment depends on the subject's responses, a greater degree of differential veracity might be expected. The effects, which such changes in conditions have on the veracity of the self-report in general, and in specific areas, need investigation. While studies on "deception" are relevant to this problem, considerations in the next section suggest that most of these can more appropriately be considered as studies of validity rather than as studies of veracity. One major exception might be noted, the studies of deceit by Hartshorne and May (7).

The high degree of relationship obtained in the present study was between the reports of normal subjects and observers. If in studies conducted in clinics and hospitals the reliability of the observers' reports and the veracity coeffiicients approximate the reliability of the patients' records, it would appear that the diagnostic inventories might well be taken by the spouse in terms of the patient; this would be particularly useful in the case of non-communicative patients. Also, it might be hypothesized that the extent of agreement between observers and subjects would be of prognostic significance: For example, for some kinds of disorders less agreement between spouse and patient might be expected in the case of acute breakdown than in those disorders of long standing and thus it would be a sign of favorable prognosis. In other disorders disagreement may be an unfavorable sign, indicating a high degree of functional disorganization. A relationship between veracity of response and reaction to psychological and medical therapies might also be anticipated. These hypotheses would appear to present empirical problems worthy of investigation.

3. *Corrections for "Low Veracity"*

The inventory approach has been severely criticized by a number of investigators. For example, Patterson has contended that the questionnaire ". . . appears to be a fruitless technique for the study of personality" (16, p. 9). In reply, the advocates of the inventory approach appear to maintain that it is fruitless only if the subject's responses deviate from the *pattern* truly characteristic of him, and then only if one is unable to correct for the deviation. [Cf. Meehl and Hathaway (13) for a summary of the literature in this area.]

Measures of this "deviation" have been included in such inventories as the *MMPI* and the Humm-Wadsworth Temperament Scale (1, 13). In the *MMPI* the measures of deviation include the *L, F,* and *K* scales: Studies of the high scorers on the *L* and *F* scales (8, 13) indicate that for them scores on the personality variables do not have the usual validity. In their use it is frequently assumed (at least implicitly) that they indicate the likelihood that the subject's responses deviate from the pattern of behaviors, interests, feelings, etc., truly characteristic of him, i.e., the likelihood that his item-responses lack veracity. Studies of the *K* scale (4, 5, 8, 10) indicate that the diagnostic significance of certain scale scores depends on the *K* elevation and it is generally assumed that for these the *K* score reflects the direction and amount of departure from veracity.

The assumption that the *L, F,* and *K* scales of the *MMPI* measure lack of veracity has not been subjected to adequate tests. The distinction drawn in this paper between validity and veracity suggest that although the *K, L,* and *F* scales may be adequate indicators of validity, they may not necessarily measure veracity: The *L* and *F* scores may only indicate the extent to which the usual *MMPI* interpretations are not valid. Similarly, the amount of elevation on the *K* scale may generally be unrelated to veracity while indicating for which subjects the presence of a few or many "claimed" symptoms are of diagnostic importance. Apparently related to this problem are the studies on "deliberate deception", such as those conducted by Meehl and Hathaway (13) and Gough (4, 5) with the *MMPI*. In these studies subjects have usually been asked to take the test under normal conditions and again with instructions to "fake." As Meehl has pointed out (12), no assumption of veracity need be made for valid protocols and hence these studies show the ability of the *L, F,* and *K* scales to measure validity and not their ability to measure lack of veracity.

The preceding considerations and the results of the present study suggest that the extent to which Patterson's contention is valid depends *not* on the inability of the inventory to obtain adequate information about the subject's behaviors, feelings, emotions, beliefs, and his perception of these, but rather on the present inability of personality theories to assimilate the information. Put another way, it may be that present conceptions of personality do not adequately take account of the subject's item-responses even when these correspond to the antecedent reactions to which they refer.

4. *The Relationship between the Concept of Veracity and Studies on "Understanding of Others"*

On the surface the present study appears similar to a number of studies on "understanding of others" (2), that is the studies on "empathy," "social sensitivity," "diagnostic competence," etc. In these a judge typically "predicts" how a subject will respond on a personality inventory and his agreement with the subject is often used as the measure of empathy, social sensitivity, diagnostic competence, etc. In the veracity study reported here the observers were *not* asked to predict how the subjects would respond; instead they were asked to indicate whether or not the item was "true" of the subject or whether they "did not know." The latter category included attributes on which they did not have specific information, that is behaviors which they had not had sufficient opportunity to observe and those feelings, attitudes, beliefs, etc., on the subject's part of which they did not know. Thus the veracity study asks the observer to indicate what *information* he has on the subject while a study on "understanding of others" asks the judge to *predict* other reactions on the basis of the information he has.

The "understanding of others" studies typically do not differentiate between the items which the judges can answer directly in terms of the information which they have and those for which they must predict or guess the answers. In the present study the observers considered that they had information on 96 per cent of the Behavioral Inventory items and 93 per cent of the *MMPI* items, hence this leaves only 4 and 7 per cent of the items respectively for them to predict or guess the answers. If this had been an "empathy" study in which all items were answered, a judge's accuracy score (percentage of agreement) would be high, not because of ability to "predict" or "guess," but because amount of information available made up such a large part of the score. It would appear that the accuracy score can not be considered as a measure of ability to "predict" the reactions of others unless amount of information is taken into consideration.

The results presented earlier indicate that there are large individual differences in ability to observe or to get information and that in this sample the amount of information is not related to length of observation (marriage). The negative relationship reported between accuracy and number of items answered (relationships which appear to increase as the reactions are less observable), indicate that if the observers had answered the "Don't know" items as either "True" or "False," on these items there would have been little agreement between subjects and observers. In fact the question may be raised as to whether there would be greater than chance agreement between the judges, predictions on these items and the subjects' responses. The problem is raised as to what extent "empathy," "social sensitivity," and "diagnostic competence" are just other names for high observational achievement and for ability to elicit information.

In a number of the "understanding of others" studies, length of the observational period has been held constant; however, it cannot be assumed that this procedure equated information for the judges. Differential achievement or information getting in the same observational situation is evident in the performance of any group of students given an examination over the lecture material covered during the preceding hour. It may be just as unwise to assume that observers or judges in the "understanding of others" studies are equal in the ability to elicit or obtain the same amount of information for a constant time interval as it would be to make this assumption for a group of students.

Cronbach (2) has analyzed the accuracy score in somewhat different terms. His analysis is in terms of the elevations and dispersions of the scores of subjects and observers irrespective of the relationship of the item-responses to the "true state of affairs." Since he is only concerned with the correspondence of the observer's report to that of the subject, he does not differentiate between responses based directly upon observational information and those based purely upon prediction or guess. In studies such as the present one in which observers answer items on the basis of observational information, his analysis does not appear applicable; however, this in no way reflects upon the value of his concepts for "understanding of others" studies in which agreements based directly upon information have been deleted.

H. SUMMARY

The concern of the present paper has been the ecological veracity of the self-report, i.e., the extent of correspondence between the self-report and the antecedent conditions and reactions to which the item-responses refer. In estimating this correspondence in the present study, two measures were used: (*a*) the percentage of agreement between subject and observer on items both answered, and (*b*) the veracity coefficient, the correlation between scale scores on the subjects' and observers' reports.

The subjects of the study were 43 married male college students. The observers, who had ample opportunity to observe the subjects' reactions in a large number of representative life situations, were the subjects' wives. Testing conditions were such that the subjects were unaware that the accuracy of their reports was being investigated. The subjects' reports consisted of their responses to items of three *MMPI* scales, *MF, SS,* and *Si,* and of three comparable scales from the Behavioral Inventory (17). The observers answered the same items, couched in the third person, in terms of their husbands. The instructions specified that they should omit items of behavior which they had not observed and items referring to feelings, attitudes, etc., on the husband's part of which they did not know.

The analysis of the data indicate that in most instances there was approximately as high agreement between observers and subjects as the reliability of the self-reports and of the observational records permit. Under such circumstances as these it is to be expected that only small and insignificant amounts of disagreement between subjects and observers would be explained unless at the same time major sources of unreliability of the scores are also explained. However, it seemed desirable to attempt such an analysis for two reasons: The split-half reliabilities might be underestimates of reliability and, secondly, it might be possible to explain the disagreement in the one or two exceptions to the noted tendency. In the analysis one major variable emerged as a possible explanation of disagreement, the tendency for percentage of disagreements to increase as the observers' cautiousness decreased, that is, as the observers made less frequent use of the "don't know" category. This tendency was especially marked for items rated as "less observable" by independent judges. These and other considerations suggest that the veracity measures may be under-estimates of the true veracity of the self-report in those areas less open to observation.

In the data analysis two concepts of deviancy were developed: Positive deviancy, the tendency for individuals to engage in unusual behaviors and to express unusual interests, feelings, emotions, and other assumed latent structures; and negative deviancy, the lack of usual behaviors and the lack of expressions of usual interests, feelings, emotions, etc.

The suggestion was made that the inventory procedure may offer an economical means of determining not only the variety of manifestations of psychological variables but also the frequency with which different means are used to satisfy different needs or motivations and the frequency with which different means are used to express emotions and other assumed latent structures. Veracity checks are conceived to be an integral part of the procedure.

Among the other issues considered were the following: Circumstances in which the veracity of the self-report should be checked. The use of the method in areas not covered in the present research. The question as to whether the questionnaire method is a suitable means of studying personality. And the relationship of veracity studies to investigations on "understanding of others."

REFERENCES

1. CRONBACH, L. J. Essentials of Psychological Testing. New York: Harper, 1949.
2. ———. Processes affecting scores on "understanding of other" and "assumed similarity." *Psychol. Bull.,* 1955, **52**, 177-193.
3. GOODENOUGH, F. L. Mental Testing—Its History, Principles, and Applications. New York: Rinehart, 1949.
4. GOUGH, H. G. Simulated patterns on the Minnesota multiphasic personality inventory. *J. Abn. & Soc. Psychol.,* 1947, **42**, 221-225.
5. ———. The F minus K dissimulation index on the Minnesota multiphasic personality inventory. *J. Consult. Psychol.,* 1950, **14**, 408-413.
6. GUILFORD, J. P. Psychometric Methods. New York: McGraw-Hill, 1936.
7. HARTSHORNE, H., & MAY, M. A. *Studies in Deceit.* New York: Macmillan, 1928.
8. HATHAWAY, S. R., & MCKINLEY, J. C. Manual for the Minnesota Multiphasic Personality Inventory (Rev. ed.). Minneapolis: 1951.
9. HOPPER, H. E., & PINNEAU, S. R. Frequency of regurgitation in infancy as related to the amount of stimulation received from the mother. *Child Devel.,* 1957, **28**, 229-235.
10. MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The Minnesota multiphasic personality inventory: VI. The K Scale. *J. Consult. Psychol.,* 1948, **12**, 20-31.
11. MCNEMAR, Q. Psychological Statistics. New York: Wiley, 1949.
12. MEEHL, P. E. The dynamics of "structured" personality tests. *J. Clin. Psychol.,* 1945, **1**, 299-303.
13. MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the Minnesota multiphasic personality inventory. *J. Appl. Psychol.,* 1946, **30**, 525-564.
14. MILTON, A., & PINNEAU, S. R. The relationship of reported overt behavior to personality dimensions. (Paper read at the Western Psychological Association, Seattle, June, 1953.)
15. ———. The relationship of reported overt behavior to social status. (Paper read at the Western Psychological Association, Long Beach, 1954.)

16. PATTERSON, C. H. The relationship of Bernreuter scores to parent behavior, child behavior, urban-rural residence, and other background factors in 100 normal adult parents. *J. Soc. Psychol.,* 1946, **24**, 3-49.

17. PINNEAU, S. R. & MILTON, A. The Multi-Dimensional Behavioral Inventory. Berkeley: 1952.

18. ———. The relationship of reported overt behavior to psychological variables. (Paper read at the Western Psychological Association, May, 1954.)

Institute of Human Development
University of California
2739 Bancroft Way
Berkeley 4, California

Department of Psychology
University of Colorado
Boulder, Colorado

The Journal of Genetic Psychology, 1958, **93**, 277-286.

CULTURAL PRESSURE AND *WAIS* SCATTER IN A TRADITIONAL JEWISH SETTING*

Psychological Center, Graduate School of Education, Yeshiva University

BORIS M. LEVINSON

A. INTRODUCTION

It is the purpose of this research to demonstrate that frequently found differences, however reliable and valid they are, between the means of the subtests of the *WAIS*, are not necessarily the result of pathology, but may occur because of subcultural differences.

It is the writer's hypothesis that the current emphasis in certain American subcultures on the development of language abilities and the neglect of non-verbal abilities effects the scatter on the Wechsler Adult Intelligence Scale. As a case in point, the *WAIS* scatter in a traditional Jewish setting is studied.

B. PREVIOUS RESEARCH

A review of literature pertaining to this problem does not reveal any investigation in the exact area of this study. There are, however, related studies. These may be divided into (*a*) the effect of socioeconomic background or deviant culture on intellectual performance, (*b*) research on adult Jewish intelligence, and (*c*) studies of psychometric scatter.

Studies show that in other cultures, the pattern of successes and failures on verbal and performance tests vary in relation to cultural needs. Furthermore, the discrepancy between verbal and performance scores indicate the cultural pressures upon the individuals. An investigation made of the performance on the Goodenough Draw-A-Man Test of various Indian tribes, as well as American minorities, indicated that the score achieved depends upon the type of cultural experience of the subject (4). The California Mental Maturity Scale was administered to freshmen at the University of Ceylon. The resulting scores of a very high verbal ability and low performance ability were interpreted in terms of Ceylonese cultural pressures upon the individual (14, 15). Eells, *et al.* (2) have shown that items in intelligence tests have "cultural differentials" which discriminate in favor of children who come from a higher socioeconomic level. It has also been the writer's general clinical experience that children from a higher socioeconomic

---

*Received in the Editorial Office on October 17, 1957.

level perform better than children from low socioeconomic levels on both performance and verbal tests.

Held (7) compared the performance on the 1938 American Council Psychological Examination of 206 male freshmen, 103 of whom were Jewish. These students were attending either the College of Liberal Arts or the School of Business Administration of the University of Pittsburgh. The Jewish boys did better in the linguistic area and the Gentile boys did better in the quantitative area. However, the difference in either intra or inter group performance was not statistically significant.

Wechsler (17), on the basis of his studies of Wechsler-Bellevue Test at Bellevue Hospital, presents characteristic configurations of subtests differentiating organic, functional, neurotic, and psychopathic disorders. Rapaport (13), who studied the scatter of subtests of Wechsler-Bellevue of 217 psychiatric patients at the Menninger Clinic, also describes patterns of scores for various psychiatric disorders.

The implications of these theories regarding the meaning of scatter on the Wechsler-Bellevue and inferentially on *WAIS* have been criticized on the grounds of (6, 12): (*a*) the low reliability of the subtests for individual diagnosis, (*b*) the high intercorrelation between the subtests, (*c*) the scatter obtained might have a different meaning at various age levels, (*d*) the failure of the pattern to appear when a new group, bearing the same nosological label, is examined.

However, few psychologists have questioned the validity of pathological implications when reliable differences in the pattern appear (9). Little heed has been paid to the fact that these differences may be due to experiential and cultural artifacts which have no bearing upon pathology. Furthermore, what may appear to be an index of pathology in one context, may actually be a sign of strength in another. In the writer's experience, this is particularly true in the case of an individual who is a product of a deviant subculture and who comes in or is referred for diagnostic study. The credulous clinician throws all his critical faculties and clinical acumen overboard and analyzes the pattern in accordance with Rapaport's (13) and Wechsler's theses (17). The justification being, that since the individual came in for treatment, "the probabilities of the difference being genuine are much higher" (16, p. 538).

C. Problems

The problems considered in this study are as follows:

1. What influence do subcultural values emphasizing language abilities have on verbal and performance *WAIS* scores?

2. What influence do these subcultural values have on the *WAIS* scatter? The group chosen for study was selected from Yeshiva University students.

D. PROCEDURE

An attempt was made to select only those Yeshiva students who could meet the following criteria and thus be considered typical products of the cultural mold exerted by traditional Jewish values:

(*a*) All were subjected to the traditional Jewish values and were survivors of a rigorously selective educational process. All attended traditional, all-day Yeshiva elementary schools and high schools.

(*b*) All had a good command of English. (While the writer has no data on bilingualism, he doubts very much whether it had any effect on the verbal scores, in view of the fact that studies indicate no correlation between bilingualism and verbal intelligence for a much younger Jewish group: children ages 9 to 14 (11).)

(*c*) All of them were normal individuals. They were not known to have personality difficulties which would have required psychotherapy.

The original sample consisted of the *WAIS* protocols of 56 male Yeshiva University students. We later added the records of eight male Yeshiva students who came in for vocational guidance, did not present any pathology, and whose *WAIS* records did not deviate from the average of the rest of the sample. Our sample finally consisted of 64 subjects, classified as follows: (*a*) six senior Yeshiva High School students, (*b*) 31 Yeshiva College students, (*c*) 27 graduate students. Among them were four ordained rabbis. Every graduate student was an alumnus of Yeshiva College.

The writer believes that the sample secured was fairly representative of the Yeshiva population. If anything, it would tend towards the median in emotional adjustment, since the subjects were all volunteers and would be approached by the examiners on the basis that their records would serve as a contribution to science. A few students, thus approached, refused to coöperate. It is thus likely that a subject who knew that he may deviate, would refuse to participate in the study.

The subjects were examined either in the dormitory or at home. In all cases, an attempt was made by the examiners to motivate the subjects and to avoid fatigue factors.

The writer wishes to add that, since the completion of the study, additional subjects were examined and, in every case, the indicated trends were confirmed.

The mean age of the group was 21.43 with a *SD* of 8.85; median age 21 years, eight months; range 16 to 31 years.

E. RESULTS

The verbal *IQ* was 125.59 with a *SD* of 10.49: median *IQ* 127; range 101 to 142. The performance *IQ* was 105.30 with a *SD* of 10.85; median *IQ* 104; range 83 to 134. The full scale *IQ* was 117.86 with a *SD* of 8.85; median *IQ* 119; range 98 to 139. The difference between the verbal and performance *IQ*s was 20.29; median difference 21; the range in differences from —2 to 47. *SD* of the difference was 1.88 and the critical ratio was 10.79. This is at the .01 level of confidence. In 62 cases (97 per cent) the verbal *IQ* was higher than the performance *IQ*.

Other differences found are suggestive, even though, in view of the small sample and possible other artifacts, no definite conclusions may be drawn. The range of differences between the verbal and performance *IQ*s for the graduate students ($N=27$) was 5 to 47; mean difference 25.85; median difference 26. For the rabbis ($N=4$), range of differences between verbal and performance was 24-41; mean 33.50; median 34.5. This would seem to give further confirmation to the hypothesis that there is a correlation between exposure to traditional Jewish cultural values and differences in scores on verbal and performance *WAIS* scales.

Note that, although these subjects were of superior verbal intelligence, their performance ability only slightly surpassed the general population level. That is, we are not dealing with individuals who have harmoniously developed abilities, but rather with men whose verbal abilities appeared to have been subjected to verbal disciplines to the relative neglect of performance abilities.

As was noted, in some cases, the *WAIS* differences between verbal and performance tests were as high as 40 points and, in other cases, as low as minus two points. It would be interesting to follow up these cases to find out why there were differences in individuals who were apparently influenced by the same cultural pressures. We might find *specific* variables to account for these differences which could be studied more intensively.

Another problem which may arise is whether the relationship commonly found between the verbal and performance scales also holds true for this population. Unfortunately, Wechsler (18) does not cite intercorrelations between the various scales for an age group similar to the Yeshiva sample. However, he does mention intercorrelations for ages 18-19 for 200 subjects consisting of 100 males and 100 females, and for ages 25-34 for 300 subjects consisting of 150 males and 150 females. In comparing the groups, the writer selected the lowest intercorrelations and the least number of subjects given by Wechsler, so that, if the *CR* is found to be significant, it would be even more significant if the intercorrelation for the general population sample similar to the Yeshiva group were known.

TABLE 1
THE CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE INTERCORRELATIONS OF THE *WAIS* SCALES OF YESHIVA STUDENTS AND THE GENERAL POPULATION (18)

Tests	Yeshiva N	Yeshiva r	General population N	General population r	SE*	CR
Verbal vs. Performance	64	.31	200	.77	0.147	4.78
Verbal vs. Full Scale	64	.78	200	.95	0.147	5.32
Performance vs. Full Scale	64	.67	200	.92	0.147	5.32

**SE* of the difference between two *z* coefficients (3, p. 239).

Table 1 presents the critical ratios of the differences between the intercorrelations of the *WAIS* scales of the Yeshiva students and the general population. We note that all of our differences are at the .01 level of confidence. While it is true, of course, that the *r*'s are lower in a group of restricted variability (3), the difference in *r*'s, however, is too great to be accounted for by that factor alone. This indicates that a much higher relationship exists between verbal and performance ability in the general population than in our group. In other words, when a Yeshiva student has a high verbal score, it does not necessarily carry the implication that he will have a high performance score.

The *WAIS* has tables for scaled scores which enable one to compare the Yeshiva student's performance on each of the tests with the average for men of his own age, in the normative group. The *WAIS* test subtest scores are thus adjusted to age differences. Since these scores (18) have the same meaning statistically at every age level, they may be combined into a scale. This was done in our sample. Every raw score was converted into a scaled score.

The medians, means, and *SD*s for the scaled score equivalents of the subtests of the verbal and performance parts of the Wechsler Adult Scale are presented in Table 2.

An analysis of Table 2 indicates a wide range in the means of the subtests of the verbal and performance parts of the *WAIS*. The means range from 8.84 to 15.64. Note the sharp drop in scores on the performance tests, also the fact that coming from the same subculture does not eliminate individual differences. As a matter of fact, it may serve to exacerbate them. If we were to accept commonly held clinical assumptions regarding psychometric scatter, these might be interpreted as indicating fair performance in visual motor tests requiring eye-hand coördination, fair graphomotor performance, and fair ability in handling three-dimensional space relationships (13). If, on the other hand, we accept the factorial structure of the *WAIS* as valid for our group, we may say that the Yeshiva students have excellent verbal comprehension and memory and average perceptual organization (1).

TABLE 2

MEDIANS, MEANS, *SD*s, AND SCORE RANGES OF THE SCALED SCORE EQUIVALENTS OF THE SUBTESTS OF THE *WAIS* OBTAINED BY YESHIVA STUDENTS

	Median	Mean	SD	Range
	Verbal			
Information	16	15.64	1.95	9 — 19
Comprehension	16	15.49	2.43	9 — 19
Arithmetic	14	13.84	2.50	9 — 19
Similarities	13	13.20	2.58	8 — 19
Digit Span	13	13.39	2.62	8 — 19
Vocabulary	14	14.48	2.38	10 — 19
	Performance			
Digit Symbol	12	12.02	2.88	6 — 18
Picture Completion	12	11.75	2.40	7 — 18
Block Design	11	11.02	2.60	5 — 17
Picture Arrangement	10	10.30	2.09	6 — 15
Object Assembly	9	8.84	2.65	2 — 17

In order to find out whether the differences between the means shown above are valid, the critical ratio of the differences between the means of the subtests of the Yeshiva students were computed.

Table 3 presents critical ratios of the differences between the means of the subtests of the Wechsler Adult Intelligence Scale of Yeshiva University students. Note that of the 55 critical ratios possible, five or 9.1 per cent are significant at the .02 level of confidence and 41 or 74.5 per cent are significant at the .01 level of confidence.

The students stated that the performance items appeared childish and unimportant. While they tried to achieve a good score whenever possible, the verbal items were more of a challenge. In discussing the digit span test, quite a few of them indicated that they had visualized the digits. This is similar to the experience of some of these students who, in studying the Talmud, remember page location of a passage. They said that an inferior score on a performance test was not as damaging to their self-esteem as a poor mark on the verbal tests.

F. DISCUSSION

The atypical psychometric pattern of Yeshiva University students may be explained on several grounds:

First, undoubtedly a somewhat similar psychometric pattern may be found in other institutions of higher learning where great emphasis is placed upon

TABLE 3

CRITICAL RATIOS BETWEEN THE MEANS OF THE SUBTESTS OF THE WECHSLER ADULT INTELLIGENCE SCALE OF YESHIVA UNIVERSITY STUDENTS

	Information	Comprehension	Arithmetic	Similarities	Digit Span	Vocabulary	Digit Symbol	Pict. Comp.	Block Design	Pict. Arrang.
Verbal										
Comprehension	0.39									
Arithmetic	4.62**	3.84**								
Similarities	6.10**	5.20**	1.45							
Digit Span	5.49**	4.77**	1.00	0.42						
Vocabulary	3.05**	2.40*	1.49	2.97*	2.42*					
Performance										
Digit Symbol	8.42**	7.54**	3.87**	2.46*	2.85*	5.23**				
Pict. Completion	10.24**	8.91**	4.86**	3.30**	3.64**	6.50**	0.57			
Block Design	11.27**	10.16**	6.27**	4.84**	5.15**	7.69**	2.04	1.62		
Pict. Arrang.	15.26**	13.30**	8.63**	7.07**	7.54**	10.45**	3.82**	3.63**	1.71	
Obj. Assembly	16.59**	15.11**	11.11**	9.69**	9.89**	12.53**	6.49**	6.47**	4.64**	3.48**

*Significant at the .02 level of confidence.
**Significant at the .01 level of confidence.

selection on the basis of high verbal aptitude and where students come from a subculture with this emphasis. Indeed, Wechsler noted that individuals of superior intelligence do better on verbal (17) than on performance tests. Parenthetically speaking, there is no a priori reason why a person of superior intelligence should have a higher verbal than performance ability, provided, of course, that both verbal and performance items are equally valid tests of intelligence. An essential factor which is being overlooked is the subculture from which the individual comes with its emphasis on verbal skills, premium on verbal accomplishment, and high standards for vocational achievement.

Second, the writer believes that the general cultural pressure has been exacerbated for this group. The subjects have been subjected, since their early school days, to a curriculum which greatly emphasized verbal knowledge, rote memory, verbal concept formation, abstract ideas, to the general neglect of performance arts. Examinations in the Talmud, for example, are oral and emphasize the detailed memorization and understanding of tracts. A differentiation of intellectual abilities has thus occurred. These cultural forces have also brought about different attitudes and self concepts regarding various intellectual tasks. Failure on a performance item does not carry the same ego deflating connotations as failures on a verbal task. It thus appears that the greater the premium placed on verbal ability in a subculture, the greater will be the disparity between verbal and performance *WAIS* scores.

This raises certain implications as far as the *WAIS* is concerned. What abilities, if any, are measured by the *WAIS* at this cultural level? What is the meaning of the full scale mean *IQ* of 117.86 for subjects who are survivors of a preschool and elementary school population whose average *IQ*s varied from 114.88 (8) to 121.3 (10)?

Furthermore, what does a performance *IQ* of 105.30 mean for this group? Is it not possible that the performance subtests measure trivial functions of little or no utility in Jewish traditional culture? Finally, is the verbal *IQ* of 125.59 representative of the intelligence of the Yeshiva University students? The writer feels that their intelligence was not adequately evaluated in view of the fact that the ceiling on some subtests of the verbal scale was not high enough for many of the subjects. What is even of greater moment, many intellectual abilities that are considered important in this group are not measured by the *WAIS*. The inference that he draws is that the validity of the *WAIS* as an index of intellectual ability for subjects whose subculture stresses the development of language abilities, should be re-examined. This should be done in terms of criteria which are considered important in that subculture.

The writer, however, does consider tests such as *WAIS* valuable tools in the study of American subcultures. These may be used to find out how cultural pressures have modified one's reactions to stimuli (5). This may give a key to the characteristic ways in which individuals meet their life's problems, thus affording a cue to a person's perception of self and his rôle in the world.

Finally, a certain scatter, repeatedly established on the same subcultural group, may give a clue to cultural stress, aiding in the comparison of subcultures and the ego values of individuals who live in those subcultures. Indeed, when a person who comes in for psychodiagnosis has a certain scattergram, it may indicate the cultural forces that impinged upon him and thus show that certain personality traits are normal *for him*. This understanding may serve as a touchstone in helping the clinician to distinguish between psychometric patterns due to pathology and those due to cultural pressures.

Research will have to determine, of course, exactly what each *WAIS* subtest measures in various subcultures. Can tests such as *WAIS* lead towards the finding of valid nationality differences?

An important implication of the above is that a person who may appear deviant in terms of his psychometric scatter as far as the general American culture is concerned, may, nevertheless, be considered as the highest product of his own subculture. He may be honored and revered for exhibiting these deviant traits and thus be a well adjusted individual within his own group.

G. SUMMARY

It was hypothesized that the emphasis in certain American subcultures on verbal accomplishment will lead to a deviant performance on *WAIS*. To verify this thesis, a sample of 64 *WAIS* scores of Yeshiva University students who have been subjected to traditional Jewish cultural values, was studied. Their mean performance *IQ* was 105.30 with a *SD* of 10.85. The mean verbal *IQ* was 125.59 with a *SD* of 10.49. The mean full scale *IQ* was 117.86 with a *SD* of 8.85. The difference between the verbal and performance *IQ*s was at the .01 level of confidence. The means of the subtests ranged from 8.84 to 15.64. Out of 55 critical ratios of differences between means of the subtests of the *WAIS,* 5 or 9.1 per cent were at the .02 level of confidence and 42 or 74.5 per cent were at the .01 level of confidence. The inference drawn is that cultural influence of traditional Jewish values have brought about this psychometric pattern and that any American subculture which emphasizes verbal abilities will bring about a somewhat similar deviant pattern. Furthermore, the greater the premium placed on verbal achievement

in a certain subculture, the greater will be the disparity between verbal and performance *WAIS* scores. The psychologist, therefore, need not necessarily attach any pathological significance to such a scatter. Research in subcultural differences in *WAIS* patterns is suggested.

REFERENCES

1. COHEN, J. The factorial structure of the *WAIS* between early adulthood and old age. *J. Consult. Psychol.,* 1957, **21**, 283-290.
2. EELLS, K., DAVIS, A., HAVIGHURST, R. J., HERRICK, V. E. & TYLER, R. W. Intelligence and Cultural Differences. Chicago: Univ. Chicago Press, 1951.
3. GARRETT, H. E. Statistics in Psychology and Education. New York: Longmans, Green, 1953.
4. GOODENOUGH, F. L., & HARRIS, D. B. Studies in the psychology of children's drawings: II. 1928-1949. *Psychol. Bull.,* 1950, **47**, 369-433.
5. GREEN, T. C. Individual response to cultural determinants. *J. Educ. Sociol.,* 1953, **26**, 392-399.
6. GUERTIN, W. H., FRANK, G. H., & RABIN, A. I. Research with the Wechsler-Bellevue Intelligence Scale, 1950-1955. *Psychol. Bull.,* 1956, **53**, 235-257.
7. HELD, O. C. A comparative study of the performance of Jewish and Gentile college students on the American Council Psychological Examination. *J. Soc. Psychol.,* 1941, **13**, 407-411.
8. LEVINSON, B. M. The intelligence of applicants to Yeshiva day schools. *Jew. Soc. Stud.,* 1957, **19**, 129-140.
9. MCNEMAR, Q. O. *WAIS* difference scores. *J. Consult. Psychol.,* 1957, **21**, 239-240.
10. NARDI, N. Studies in intelligence of Jewish Children. *Jewish Educ.,* 1948, **19**, No. 3, 41-50.
11. PINTNER, R., & ARSENIAN, S. The relation of bilingualism to verbal intelligence and to school adjustment. *J. Ed. Res.,* 1937, **31**, 255-263.
12. RABIN, A. I., & GUERTIN, W. H. Research with the Wechsler-Bellevue Test: 1945-1950. *Psychol. Bull.,* 1951, **48**, 211-248.
13. RAPAPORT, D., GILL, M., & SCHAEFER, R. Diagnostic psychological testing, Vol. 1. Chicago: Year Book Publishers, 1945.
14. STRAUSS, M. Mental ability and cultural needs: A psychocultural interpretation of the intelligence test performance of Ceylon University entrants. *Amer. Soc. Rev.,* 1951, **16**, 371-375.
15. STRAUSS, M. Subcultural variation in Ceylonese mental ability: A study of national character. *J. Soc. Psychol.,* 1954, **39**, 129-141.
16. THORNDIKE, R. L., & HAGEN, E. Measurement and Evaluation in Psychology and Education. New York: Wiley, 1955.
17. WECHSLER, D. The Measurement of Adult Intelligence. (3rd ed.) Baltimore: Williams & Wilkins, 1944.
18. ———. Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corporation, 1955.

Psychological Center
Graduate School of Education
Yeshiva University
110 West 57th Street
New York 19, New York

The Journal of Genetic Psychology, 1958, **93**, 287-290.

PREIRRADIATION FATIGUE AS A FACTOR IN THE PREVENTION OF IRRADIATION DEATHS IN RATS*

Radiobiological Laboratory, University of Texas

W. LYNN BROWN AND ROBERT K. WHITE[1]

A. PROBLEM

Kimeldorf and Jones (4) found that exhaustive fatigue by swimming increased the mortality of rats if swimming followed irradiation, but no changes in mortality rates were found as a consequence of exercise prior to irradiation. In a preliminary study on fatigue and irradiation in this laboratory, using *S*s which had been fatigued by extensive swimming, we observed that some of the *S*s which were fatigued *immediately* (within 30 seconds) prior to irradiation survived what had been shown to be an $LD_{100/10}$ dose of gamma irradiation (unpublished data).

The experiment presented here was designed to test the hypotheses suggested by the above observation that fatigue induced immediately prior to irradiation may minimize the lethality of irradiation even though it is decremental postirradiation as was found by Kimeldorf and Jones.

B. PROCEDURE

Ten five-month-old female Sprague-Dawley rats were used as subjects. Using the Kimeldorf and Jones technique they were given, prior to irradiation, 10 successive days of swimming in a 24-gallon cylindrical metal can filled with fresh water to a constant depth of 18 inches. The cans were cleaned and refilled daily. The water temperature was 20° ± 1.5° C. A 10-gm. weight was clipped to the chest hair of the *S* during each swimming period. Only one animal was placed in the container at a time. The *S*s swam until they sank 18 inches below the surface of the can and remained below the surface for an estimated 20 seconds before they were retrieved. Swimming exercise was given daily for 10 days between 8-10 A.M. Daily feeding was between 1-3 P.M.

---

*Receiv d in the Editorial Office on October 21, 1957.

[1] This work was conducted at the Radiobiological Laboratory of The University of Texas and the United States Air Force, Austin, Texas and was supported (in part) with funds provided under Contract AF 41(657)-149 and the School of Aviation Medicine.

Following the tenth period of swimming the *S*s were divided into two treatment groups matched for weight. Group I was fatigued on the eleventh day of the experiment and immediately irradiated following removal from a swimming container. Group II was irradiated on the same day but was not fatigued. Neither group had swimming exercise postirradiation. Both groups were irradiated with 1,200 *r* cobalt^{60} at 600 *r* per minute in the high-level Co^{60} facility (1) at this laboratory. *S*s were placed in individual containers and rotated during irradiation to produce radiation symmetry. The $LD_{50/30}$ for the rat in this facility is 950 *r* at 500 *r* per minute, and $LD_{96/30}$ is 1,100 *r* at the same rate (7) with 3½-month-old Sprague-Dawley rats as experimental animals. The *S*s were not fed the day of irradiation.

All *S*s in Group II died within 10 days postirradiation. All in Group I lived 30 days postirradiation (see Figure 1). A Mann-Whitney (6, pp. 116-127) test applied to this data gave a *U* value of 0. With $N_1 = N_2 = 5$ this value is significant beyond the .01 level for a two-tailed test of significance.

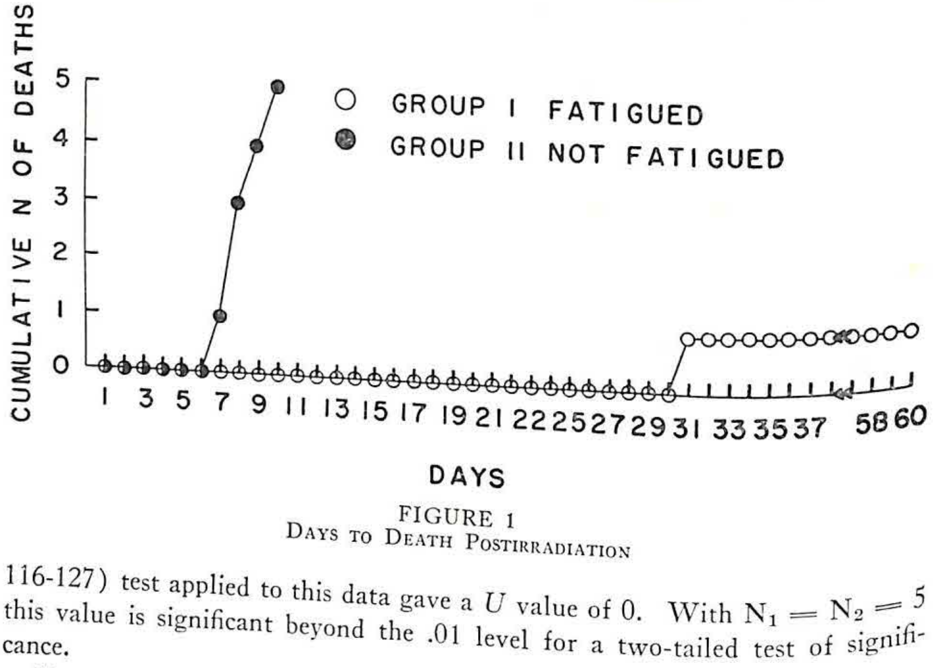

FIGURE 1
Days to Death Postirradiation

Since all the *S*s in Group I lived one month postirradiation and all in Group II were dead within 10 days, these results confirm the hypothesis that fatigue at the time of irradiation has a protective effect. Apparently the *S*s which were in a state of fatigue at the time of irradiation were protected from the lethal dose of irradiation. We assume that the changes produced in the organism by swimming until fatigued immediately prior to irradiation served as a protective mechanism against a lethal dose of irradiation. Theoretical explanations of these results can take many different directions, since the physiological state of the organism, especially at the tissue or subcellular level, produced by this fatiguing procedure is undoubtedly a highly complex one.

Wells, Blake, and Van Fossan (8) have pointed out that one of the characteristics of the "exhausting fatigue" state is ". . . the accumulation of acid metabolites or the unsatisfied demand for adequate amounts of oxygen supply . . ." (p. 4). The increased oxygen debt of the tissue is represented by the accumulation of lactic acid (2). If the anaerobic condition of the tissue which results in this extreme accumulation of lactic acid may be considered as a form of anoxia, then the facilitory effect of fatigue at the time of irradiation could be explained by referring to the already proven beneficial effect of anoxia. Patt (5) in his review concluded that ". . . the radiation dose required to induce a comparable degree of change in the absence or near absence of oxygen is generally increased by a factor of 2 to 3 . . ." (p. 46). From the present experiment it appears that anoxia might be a factor in producing the facilitory effects of exhaustive fatigue at the time of irradiation.

C. SUMMARY

Ten female Sprague-Dawley rats were given 10 successive days of swimming until fatigued each day. Five of the *S*s (Group I) were irradiated on the eleventh day with 1,200 *r* of cobalt^{60} immediately following exhaustive fatigue produced by swimming and five (Group II) were exposed to the same radiation without swimming exercises. All *S*s in Group I were alive one month postirradiation while all *S*s in Group II were dead within 10 days postirradiation. It was assumed that exhaustive fatigue, or anoxia produced by fatigue, functioned as a protective mechanism against a lethal dose of irradiation for Group I.

REFERENCES

1. AUXIER, J. A. A high level Co^{60} irradiation facility. *Sch. Avia. Med., USAF*, Report No. 55-40, June, 1955.
2. BEST, C. H., & TAYLOR, N. B. The Physiological Basis of Medical Practice. Baltimore: Williams & Wilkins, 1945.
3. DOWDY, A. H., BENNETT, L. R., & CHASTAIN, S. M. Protective action of anoxic anoxia against total-body roentgen irradiation in mammals. *Radiology*, 1950, **55**, 879-889.
4. KIMELDORF, D. J., & JONES, D. C. The relationship of radiation dose to lethality among exercised animals exposed to roentgen rays. *Amer. J. Physiol.*, 1951, **167**, 626-632.
5. PATT, H. M. Protective mechanisms in ionizing radiation injury. *Physiol. Rev.*, 1953, **33**, 35-76.

6. SEIGEL, S. Nonparametric Statistics. New York: McGraw-Hill, 1956.
7. TATSCH, R. E., TOURVILLE, J. C., & LOGIE, L. C. Using Co^{60} to determine the $LD_{50//30}$ for the rat. *Sch. Aviat. Med., USAF*, Report No. 57-89 (in press).
8. WELLS, J. G., BLAKE, B., & VAN FOSSAN, D. D. Lactic acid accumulation as a factor in determining work capacity. *Sch. Aviat. Med., USAF*, Report No. 56-121, 1956.

Radiobiological Laboratory
University of Texas
Austin, Texas

The Journal of Genetic Psychology, 1958, **93**, 291-301.

A STUDY OF SEVENTH GRADE CHILDREN'S READING OF COMIC BOOKS AS RELATED TO CERTAIN OTHER VARIABLES*

College of Education, Drake University

W. Paul Blakely

A. Introduction

Speculation about alleged dangers as well as possible benefits of the ubiquitous comic books to their numerous youthful readers, has been widespread and vehement. There is less empirical data than the important status and serious implications of this problem would seem to warrant. Gunnar Dybwad, Executive Director of the Child Study Association of America, speaking before the United States Senate Subcommittee Investigating Juvenile Delinquency in 1954, declared that: "The absence of any definitive studies of the effects of comic reading on children's emotions and/or behavior has been a serious handicap to us as to everyone dealing with this problem." The Senate Subcommittee itself, in its Interim Report issued in 1955, noted "with some surprise that little attention has been paid by educational and welfare agencies to the potential dangers, as well as benefits, to children presented by the growth of the comic book industry."

The purpose of the present study has been to identify among a specific group of children, subgroups whose reading of comic books falls into various categories as to amount and as to type; then to measure as well as possible among these children certain variables upon which the reading of comic books has been alleged to have a bearing; and to determine whether the subgroups differ on the average with respect to these variables as measured.

B. Related Literature

It is rather easy to find literature in which this problem is discussed in terms of clinical or case-study material, or in terms of "expert opinion." It is more difficult, as already implied, to find reports of experimental data.

Dr. Frederic Wertham, in his status as a practicing psychiatrist, has commanded a wide audience with various pronouncements condemning comic books. Citing case material on children with reading difficulties whom he has found to be "addicted" to the reading of comic books, for example, he infers a causal connection (8).

---

*Received in the Editorial Office on October 21, 1957.

Sperzel, in an experimental study involving use of comic books in the classroom, found no effects on reading ability (7). Luckiesh and Moss have cited statements of ophthalmologists that some children appear to suffer from eyestrain attributable to reading the comics, but have noted at the same time the lack of evidence available (2).

Witty has made inferences about comic reading and educational achievement in a group of Evanston-Chicago-Milwaukee children he studied, reporting that a subgroup who read the most comic books received almost the same average school marks as one who read the fewest (9). Heisler (5) studied 600 elementary and junior high school children and for each school grade level, made a number of comparisons between a group who read no comic books and one of equal or nearly equal size who read the most. Evidences of differences in school achievement, adjustment, and ownership of library-type books were neither consistent nor of a generally recognized significance level.

Concerning adjustment and behavior, Dr. Wertham (8) says that his clinical data obtained by a number of standard techniques show factors in the psychological patterns of children which require other than conventional Freudian explanations, and correspond to material in comic books. He cites as an example, an injury-to-the-eye motif.

Josette Frank (4) in 1948 reported a symposium of psychiatric and psychological opinion, concluding that there was agreement that "the radio programs, movies and comics do not in themselves *create* fears, but for certain children and under various conditions, do precipitate or stimulate anxieties lying beneath the surface ready to be awakened."

Wolfe and Fiske studied a relatively small group of children intensively and found certain unhealthy characteristics associated with reading the comic books as "fans," that is, with excessive intensity (10).

Hoult (6) studied 235 boys and girls, ages 10 through 17, who had been arrested for juvenile delinquency, in comparison with a matched group of controls. He categorized comic books as harmful, questionable, and harmless according to their general content (e.g., crime and gangsterism, animated animal cartoons), and found that his delinquent subjects reported reading more titles in the first two categories than the non-delinquents, but no more in the third.

C. The Present Study: Population and Method

The population chosen for the present study consisted of 12 class sections of seventh grade children in two public junior high schools in an Iowa city of about 100,000 population. Complete data were obtained for 323 children, except that one of the coöperating schools withheld a group of children from the achievement testing program so that achievement test scores were available on only 281.

The city is one which thrives on industry and trade. The schools in the study draw from populations which are not atypical in the city. An unselected list of parents' occupations as reported by the children in the study follows: printer, super-market manager, tool-grinder, worker in sales department of radio-manufacturing company, truckdriver, dock foreman, traveling salesman, crane operator, dishwasher (mother), painter, electrician, design engineer, roofer, and college instructor.

As a measure of comic book reading, three questionnaires were constructed, each labeled *Junior High Interest Questionnaire*, and administered at intervals of about one month beginning early in September, 1956. Among the questions on each, the child was asked to list the titles of all comic books he had read "within the past week," and in any case where he had read the same title more than once, to put the number of times just following the title. The question about comic book reading, in each case, was placed among others about various interests and activities in an effort to avoid the possibility of putting the responding child "on his guard"; that the study was particularly concerned with comic book reading was not divulged.

From the questionnaires was computed for each child his total comic book *frequency*—that is, the total number of titles reported read, multiplied by the number of times for each. (For example, having read *Donald Duck* 2 times and *Tarzan* once would constitute a *frequency* of 3.)

For each child was also computed a *comic book type score*. Each title which appeared in the most recent classification of the Cincinnati Committee on Evaluation of Comic Books was considered accordingly, as rating "No Objection," "Some Objection," or "Objectionable." For titles not on the most recent list, earlier lists were consulted. Some titles found on none of these lists were classified by the investigator, applying as best he could the criteria set forth by the Committee in its 1956 leaflet. It was impossible to classify 7.31 per cent of the reported frequency, which were ambiguous or could not be identified; these titles were ignored in computing the comic book type score. The complete list of titles reported, with their classifications, appears in the writer's doctoral dissertation, which is in the library of the State University of Iowa, and on file with University Microfilms, Ann Arbor, Michigan.

The comic book type score for each child was computed by the formula

$$T = \frac{(N \times 2) + (S \times 1) + (B \times 0)}{N + S + B} \times 10$$

where $T =$ type score; $N =$ frequency of "No Objection" titles; $S =$ frequency of "Some Objection" titles; and $B =$ frequency of "Objectionable" titles.

As a measure of "legitimate" reading, each questionnaire also asked the child to list all books of the type obtainable at a library, which he had read "within the last month," and after each, to specify where he had obtained it—public library, school library, home, etc.

As an intelligence test as independent of reading as possible, The Lorge Thorndike Intelligence Test, Non-Verbal Series, Level 4, Form *A* was administered.

The SRA Junior Inventory, Form S was chosen as a measure of various phases of adjustment. The checking of problems on the Inventory as "big" or "middle-sized" has been taken as indicative of the child's concern about his adjustment in the respective area. Checking a problem as "little" or "not a problem" might, of course, indicate conscious or unconscious reluctance to admit the problem, as well as the non-existence of the problem for the child concerned. Therefore, the scores derived by counting the number of checks in "big" and "middle-sized" boxes must be taken as indicative of the child's *perception* (quantitatively) of his problems. For each child, the total score on the inventory was thus computed, as well as the subscores "About Me and My School," "About Me and My Home," "My Health," "About Myself," "Getting Along With Other People," and "Things in General"; and a subscore for a category of items selected throughout the inventory taken to be indicative of "personal adjustment."

As a measure of behavior and behavior-related characteristics observable to others, teachers of the children filled out the Haggerty-Olson-Wickman Behavior Rating Schedules, *A* and *B*. Schedule *A* consists of a list of behavior problems in their order of frequency as reported for a group of elementary school children, with statistical weights assigned on the basis of seriousness and frequency. In Schedule *B* are five-point graphic rating scales for traits studied in relation to the behavior problems of Schedule *A* among the standardizing group; thus a high score on Schedule *B* is taken to be suggestive of behavior-problem proneness.

The Stanford Achievement Test, Advanced Battery, Form *J*, was administered during the school year as part of the regular program of the coöperating schools. Grade norm scores for the children in the study were taken as follows: reading (average of Paragraph Meaning and Sentence Meaning); average of Language and Spelling; and general school achievement status (average of Language, Spelling, Arithmetic Reasoning, Arithmetic Computation, and Social Studies).

The intelligence test, the *SRA* Junior Inventory, and Questionnaires 1 and 2 were administered by the investigator during school periods allotted him for the purpose. Questionnaire 3 and the Stanford Achievement Test were administered, and the ratings on the Haggerty-Olson-Wickman Schedules made, by coöperating school personnel.

D. PARENT INTERVIEWS

Hoping to be able to make an assessment of the truth of the children's answers to the questionnaires, the investigator visited the homes of 35 of them. The 35 were initially chosen by taking every tenth name from an alphabetized list of all children who responded to the first questionnaire. When an interview could not be completed after a second call at a home (because no one was found at home), the next name in order was substituted.

The parents were asked questions about their children's comic book and library book reading, along with others about television viewing and other activities. It soon became apparent that an estimate of their children's weekly reading of comic books is difficult for many parents to make. Likewise, much of the library book reading was found to take place in the school, and the parents were unable to report it very accurately. A product-moment correlation coefficient of 0.42 was obtained, for the 35 cases, between the average weekly frequency of comic book reading computed from each child's three questionnaires, and the parents' estimates of "number of comic books read each week." A coefficient of 0.43 was obtained between the average monthly number of library books reported read by the children, and a parent estimate of library books obtained monthly, on the average, from the public library plus "other sources."

Many of the parents interviewed obviously felt it reflected on their child-rearing to admit their children were readers of comic books. This, plus the fact that the children were asked actually to list titles of comic books and library type books read, is taken as justifying the conclusion that there is probably greater error in the parents' estimate than in the figure derived from the children's reports.

E. RESULTS OF THE STUDY

1. *Comparisons Based on Comic Book Frequency Groups*

The children participating in the study were divided into subgroups according to total reported comic book frequency, for purposes of a series of comparisons. An attempt was made to keep subgroups of maximum size possible while at the same time maintaining categories descriptive as to frequency of comic book reading. Table 1 shows the subgroupings used; it will be noted, for example, that "Light Readers" reported reading frequency of from one to six comic books during the three weeks covered by the questionnaires, or an average frequency of not more than two per week.

TABLE 1
THE SUBGROUPS ACCORDING TO TOTAL REPORTED COMIC BOOK FREQUENCY FOR THREE NON-CONSECUTIVE WEEKS

Subgroups	N*	Range	Mean	*SD*
Boys				
I. Non-readers	43 (30)	0	0	0
II. Light readers	35 (32)	1-6	3.26	1.63
III. Moderate readers	55 (46)	7-21	12.95	4.23
IV. Heavy readers	39 (34)	22 or more	44.18	27.76
Total	172 (142)		14.82	21.44
Girls				
I. Non-readers	33 (27)	0	0	0
II. Light readers	46 (44)	1-6	3.48	1.53
III. Moderate readers	54 (51)	7-21	12.07	3.96
IV. Heavy readers	18 (17)	22 or more	39.67	27.44
Total	151 (139)		10.11	15.41
Total				
I. Non-readers	76 (57)	0	0	0
II. Light readers	81 (76)	1-6	3.38	1.58
III. Moderate readers	109 (97)	7-21	12.51	4.12
IV. Heavy readers	57 (51)	22 or more	42.75	27.75
Total	323 (281)		12.62	19.01

*Figures in parentheses are reduced N's used for achievement test comparisons.

Comparisons among the subgroups so derived, with respect to the other measures, which are regarded as dependent variables, were made by the simple analysis of variance, using a .05 coefficient of risk. The comparisons are shown in Table 2.

Where the obtained F meets the chosen criterion of significance, the t-test has been applied between the various pairs of means. The following differences in reported library-book reading were thus found to be significant: Subgroups III over I and IV over I (.01 level); and IV over II (.05 level).

When the same set of analyses of variance were applied to the sexes separately, an F of 4.83 (significant at .01 level, $df = 3,168$) was obtained for boys alone in reported library-book reading; and significant t's were found in testing the following differences: Subgroups III over I, IV over II, and IV over III (.05 level); and IV over I (.01 level). Thus for boys, and for the sexes combined, reported readership of library-type books tends to increase as does frequency of comic book reading. The subgroup means for library book readership for girls clearly followed the same pattern, but not to the chosen level of significance.

TABLE 2
COMPARISONS AMONG COMIC BOOK FREQUENCY SUBGROUPS (SEXES COMBINED)

	Subgroup means					
Measure	I	II	III	IV	*df*	*F*
Lorge-Th. Non-V *IQ*	99.53	102.73	102.81	102.40	3,319	1.55
Library Book Rdg. (3 mo.)	4.00	5.14	5.73	6.68	3,319	5.12**
SRA Jr. Inv.						
Home	2.20	2.31	2.59	3.26	3,319	1.30
People	1.46	2.48	2.28	3.00	3,319	1.91
School	8.50	7.67	7.89	8.35	3,319	0.28
Self	2.50	3.24	3.28	2.95	3,319	0.62
Th. Gen.	4.59	4.98	5.31	6.18	3,319	0.99
Health	1.93	2.22	2.44	2.75	3,319	0.86
Pers. Adj.	2.45	2.91	2.89	3.07	3,319	0.28
Total	21.18	22.89	23.80	26.49	3,319	0.65
Hag-Ol-Wick						
Sch. A	11.42	7.52	11.87	13.95	3,319	1.30
Sch. B	66.34	65.35	67.60	67.07	3,319	0.29
St. Achvt.						
Reading	9.01	8.95	8.99	9.14	3,277	0.10
Lang-Sp	8.52	8.56	8.31	8.32	3,277	0.37
5 Subtests	8.47	8.45	8.34	8.28	3,277	0.27

**Significant at .01 level.

With the exception noted, the comparisons made among subgroups who differ in frequency of comic book reading show no corresponding, average differences in the measures applied.

2. *Comparisons Based on Comic Book Type Score Groups*

A second arrangement of subgroups has been made according to comic book type scores, as shown in Table 3. Again, factors of both maximum subgroup size and descriptiveness of category were considered. However, the range of type scores in Subgroup IV is necessarily wider than might be desirable, because of the considerable negative-skewedness of the type-score distribution. It will be noted that Subgroup I, "Non-readers," is the same subgroup so-used in the comparison based on frequency.

TABLE 3
THE SUBGROUPS ACCORDING TO COMIC BOOK TYPE SCORES*

Subgroup	*N***	Range	Mean	*SD*
Boys				
I. Non-readers	43 (30)	—	—	—
II. All "No Obj."	50 (44)	(All have scores of 20)		
III. Few "S.O.," "Obj."	46 (37)	16-19	17.67	1.06
IV. Most "S.O.," "Obj."	33 (31)	0-15	11.15	4.34
Total readers	129 (112)	0-20	16.91	4.19
Girls				
I. Non-readers	33 (27)	—	—	—
II. All "No Obj."	70 (66)	(All have scores of 20)		
III. Few "S.O.," "Obj."	30 (29)	16-19	17.50	1.06
IV. Most "S.O.," "Obj."	18 (17)	0-15	12.94	2.63
Total readers	118 (112)	0-20	18.29	2.75
Total				
I. Non-readers	76 (57)	—	—	—
II. All "No Obj."	120 (110)	(All have scores of 20)		
III. Few "S.O.," "Obj."	76 (66)	16-19	17.61	1.07
IV. Most "S.O.," "Obj."	51 (48)	0-15	11.98	3.92
Total readers	247 (224)	0-20	17.57	3.64

*For derivation, see text.
**Figures in parentheses are reduced *N*'s used for achievement test comparisons.

Results of the simple analysis of variance as applied to this arrangement of subgroups, with respect to the various measures applied, may be observed in Table 4. Again, the lack of significant differences is much greater than their presence. It will be noted that the tendency of comic book readers to be greater readers of library books, by their own report, at least, persists without respect to the "type" of comic books read.

When the series of analyses were applied to the type-score subgroups, sexes separately, one further difference emerged. The girls in Subgroup III score higher on problems related to "Things in General" than do those in Subgroup II (see Table 5). This category of items in *SRA* Junior Inventory, Form *S*, is rather miscellaneous; the one outstanding characteristic seems to be a lack of skills or "know-how": "I need to learn how to use the library"; "I want to learn how to dance"; "I wish I could take music lessons"; "I wonder what my real abilities are." It is probably not justified to invest too much concern in this one significant difference which appears among so many possible ones (maybe this is an error of rejection which the .05 coefficient of risk implies will occur 5 per cent of the time in the long run). However, it should be noted that girls in type-score Subgroups III or IV, or both, also have higher observed means than those in Subgroups I and/or II (not at the chosen level of significance) in *SRA* total, "People," "Self," "Health," and "Personal Adjustment," and in the Haggerty-Olson-Wickman schedules.

TABLE 4

COMPARISONS AMONG COMIC BOOK TYPE SCORE SUBGROUPS (SEXES COMBINED)

	Subgroup means					
Measure	I	II	III	IV	*df*	*F*
Lorge-Th. Non-V. *IQ*	99.53	103.39	101.54	103.00	3,319	2.00
Library Book Rdg. (3 mo.)	4.00	5.73	5.71	5.90	3,319	3.51*
SRA Jr. Inv.						
Home	2.20	2.33	3.11	2.73	3,319	1.20
People	1.46	2.08	2.79	3.14	3,319	2.59
School	8.50	7.53	8.84	7.47	3,319	0.88
Self	2.50	3.08	3.51	2.96	3,319	0.76
Th. Gen.	4.59	4.58	6.45	5.77	3,319	2.38
Health	1.93	2.32	2.57	2.55	3,319	0.65
Pers. Adj.	2.45	2.63	3.18	3.29	3,319	0.66
Total	21.18	21.93	27.26	24.61	3,319	1.26
Hag-Ol-Wick						
Sch. A	11.42	9.83	12.12	11.64	3,319	0.24
Sch. B	66.34	64.81	66.59	72.55	3,319	2.20
St. Achvt.						
Reading	9.01	8.89	9.15	9.09	3,277	0.31
Lang-Sp	8.52	8.40	8.63	7.94	3,277	1.51
5 Subtests	8.47	8.37	8.43	8.24	3,277	0.27

*Significant at .05 level.

TABLE 5

COMPARISONS AMONG GIRLS' COMIC BOOK TYPE SCORE SUBGROUPS, *SRA* JUNIOR INVENTORY, "THINGS IN GENERAL"

Subgroups	Means	Differences		
		I	II	III
I	5.36			
II	4.30	1.06		
III	7.57	2.21	3.27**	
IV	6.50	1.14	2.20	1.07

F (df = 3, 147) = 2.80*.
*Significant at .05 level.
**Significant at .01 level.

F. CONCLUSIONS

Within the limitations inhering in the method and population of the study, the following conclusions have been drawn:

1. Seventh grade children who differ in comic book reading, either as to type (as defined) or as to frequency, do not differ correspondingly in tested reading ability.

2. Such children, so grouped, do not differ correspondingly in achievement in language and spelling as measured by a standardized test.

3. Such children, so grouped, do not differ correspondingly in general school achievement as measured by an average of five subtests of a standardized test.

4. Such children, so grouped, do not differ correspondingly in behavior problem status, or in traits related to behavior problem status, as observed and rated by their teacher.

5. Such children, so grouped, are not significantly different from each other, correspondingly, in intelligence.

6. Seventh grade children who read comic books read more library books than do those who read no comic books. The difference tends to be in proportion to the frequency of comic book reading, and to exist for all "qualities" of comic book reading as defined by the Cincinnati Committee. Although the difference is not significant in the case of girls alone, their subgroup means clearly follow the same pattern as those for boys and they appear to contribute to the significance level for the combined sexes.

7. Seventh grade boys who differ in comic book reading, either as to type (as defined) or as to frequency, do not differ correspondingly in reported recognition of their own personal problems.

8. Seventh grade girls who read some comic books of the type regarded by the Cincinnati Committee as rating "Some Objection" or "Objectionable" are more troubled than are those who read entirely comic books of the "No Objection" type, by some personal problems. The difference is statistically significant in the *SRA* Inventory subdivision, "Things in General," which seems in general concerned with need for knowledge, skills, and experience. There is a suggestion, not otherwise statistically substantiated, that girls whose reading of comic books falls into these "objectionable" categories may be a somewhat less well-adjusted group than others. The point is highly inconclusive.

10. The data as analyzed provide only very meager evidence (and this in the case of girls only) that the type of criteria applied in classifying comic books according to type or "quality" identify comic books the reading of which has observable, immediate concomitants in the variables studied for seventh grade children.

Finally, it may be well to emphasize that this study does not purport to have disproved the existence of all the ills allegedly attendant upon the reading of comic books. It is a serious matter to endanger the morals, the mental health, or the normal progress of development of children. It is also a serious matter to interfere with popular access to the media of communication—certainly too serious to be undertaken casually. Those who would do either of these two things ask us to engage in a calculated risk of large proportions. Upon those proposing either type of risk it would seem incumbent to marshall justifying evidence. The results of this study for the most part *fail to support* curtailment of children's access to comic books.

REFERENCES

1. EDUCATIONAL DIGEST. Comic books and juvenile delinquency. *Educ. Dig.*, 1955, **21**, 8-10.
2. DEBOER, J. J. Using modern channels of communication: Magazines. *Elemen. Eng.*, 1950, **27**, 107-125.
3. DYBWAD, G. Statement to the Senate Subcommittee. *Child Stud.*, 1954, **31**, 38-41.
4. FRANK, J. Chills and thrills in radio, movies, and comics. *Child Stud.*, 1948, **25**, 42-48.
5. HEISLER, F. A comparison of comic book and non-comic book readers of the elementary school. *J. Educ. Res.*, 1947, **40**, 458-464.
6. HOULT, T. F. Comic books and juvenile delinquency. *Sociol. & Soc. Res.*, 1949, **33**, 279-284.
7. SPERZEL, E. Z. The effect of comic books on vocabulary growth and reading comprehension. *Elemen. Eng.*, 1948, **25**, 109-113.
8. WERTHAM, F. Seduction of the Innocent. New York: Rinehart, 1954.
9. WITTY, P. Reading the comics: A comparative study. *J. Exper. Educ.*, 1941, **10**, 105-109.
10. WOLF, K. M., & FISKE, M. The children talk about comics. *In* (Lazarsfeld, P. F., & Stanton, F. N., *Eds.*), *Communications Research 1948-1949.* New York: Harper, 1949.

College of Education
Drake University
Des Moines, Iowa

SHORT ARTICLES AND NOTES

The Journal of Genetic Psychology, 1958, **93**, 303-305.

SUBSTITUTION WITH INTERRUPTED TASKS OF DIFFERING VALENCE*

Department of Psychology, University of Wisconsin

ANITRA GORDON AND WILLARD THURLOW

A. INTRODUCTION

Tendency to complete an interrupted task has been related by Lewin to the operation of a system under tension (3, p. 184). Completion of the interrupted task presumably releases the tension of the system corresponding to it. Completion of a second task may result in a decreased tendency for the subject to complete a prior incomplete task. In this case, *substitution* is said to have taken place. Henle (1, p. 73-81; 2) reported experiments on the influence of valence on substitution, utilizing both objectively similar and dissimilar tasks. In her results it is not completely clear which valence (that of Task 1 or Task 2) was chiefly responsible for the results. The present experiment was designed to overcome these difficulties of interpretation by including suitable control conditions. These experiments, run with children, utilized objectively dissimilar tasks, and aimed to answer the following questions: (*a*) Will substitution occur when a low valence task is followed by an objectively dissimilar medium valence task? (*b*) Will substitution fail to occur when a high valence task is followed by an objectively dissimilar medium valence task?

B. PROCEDURE

The following general conditions were used: (*a*) *LM* : Low valence task, interrupted, followed by medium valence task, completed. (*b*) *L*— : Low valence task, interrupted, followed by simple questions ("Hold old are you?" etc.), taking up the same average time as *M* in the *LM* condition. Questions were kept constant for all subjects. (*c*) *HM* : High valence condition corresponding to *LM*. (*d*) *H*— : High valence condition corresponding to *L*—. In all conditions, the main interest was to see whether resumption of

---

*Received in the Editorial Office on July 11, 1957.

the interrupted task would occur. One minute was allowed for resumption. (Preliminary experimentation indicated that resumption—and completion—tended to occur within one minute if it occurred at all.)

The subjects, 32 children (age 4½-6½ years), were each run in all conditions. Conditions *LM* and *L*— (or *HM* and *H*—) were administered first; then after a pause of 15 minutes the other pair of conditions was administered. Order of conditions within pairs, and also tasks used for each valence condition were counterbalanced among subjects. The following six tasks were used: (*a*) cutting out animal, (*b*) putting together puzzle (11 pieces), (*c*) building wall, 12 blocks, (*d*) coloring in picture, (*e*) stringing 12 beads, (*f*) putting together simple form board. The six tasks were first demonstrated to the subject; then his order of preference for the tasks was determined. Tasks ranked 1 and 2 were designated high valence tasks; 3 and 4, medium valence tasks; 5 and 6, low valence tasks. Valences determined from initial rankings were used for the first pair of experimental conditions. Valences were redetermined after the 15-minute pause, and these redetermined valences were used for the second pair of conditions.

C. RESULTS

Percentage resumption (and completion) for the interrupted task was as follows: *H*—, 100; *HM*, 78; *L*—, 72; *LM*, 50. The hypothesis (derived from Henle) that *LM* should be significantly less than *L*— was confirmed (beyond .05 level, using test for correlated proportions). The hypothesis (also derived from Henle) that *HM* should not be significantly different from *H*— was not confirmed (difference significant beyond .05 level). Thus significant substitution was demonstrated when a medium valence task followed both a high valence and low valence interrupted task.

D. DISCUSSION

It should be noted that although substitution occurred significantly, a large number of the subjects did not show a substitution effect. This may well be a function of the subjects we used. The difference in subjects used must be remembered in comparing our results to those of Henle, who used college students.

This experiment is considered to be only a first step in the investigation of the influences of valence on substitution. There are a number of further questions to be answered. For instance we need to know to what extent degree of completion of the second task influences substitution.

E. Summary

"Substitution" is said to have taken place when completion of a second task results in a decreased tendency for a subject to complete a prior incomplete task. The present experiment was designed to determine the degree of substitution obtained when a medium valence task followed a high valence or low valence incompleted task. Tasks used were objectively dissimilar. Subjects were 32 children whose ages ranged from 4½ to 6½ years.

Statistically significant degrees of substitution were found. It is concluded that a medium valence task can provide significant substitution for an incomplete high or low valence task. It has also been pointed out that a large number of the children did not show a substitution effect under the conditions of this experiment.

References

1. Henle, M. An experimental investigation of dynamic and structural determinants of substitution. *Contrib. Psychol. Theo.*, 1942, **2**, No. 3.
2. ———. The influence of valence on substitution. *J. of Psychol.*, 1944, **17**, 11-19.
3. Lewin, K. A Dynamic Theory of Personality. New York: McGraw-Hill, 1935.

Department of Psychology
University of Wisconsin
Madison 6, Wisconsin

The Journal of Genetic Psychology, 1958, **93**, 307-312.

BOOKS

Now that there is a special *APA* journal completely devoted to the publication of book reviews, it is no longer necessary that other journals emphasize such publication. It has always been our conviction that book reviews are a secondary order of publication unless they carry information that is equally as important as the book. However, the publication of book titles is a very important service, and we shall continue to render that service.

In any given issue of this journal, we may continue to publish one or more book reviews, but we do not consider such publication a major function of this journal. In line with this policy, we can no longer pay for such manuscripts.

BOOKS RECENTLY PUBLISHED

Allan, W. S. Rehabilitation: A Community Challenge. New York: Wiley, 1958. Pp. 247.
Anastasi, A. Differential Psychology. 3rd Ed. New York: Macmillan, 1958. Pp. 664.
Beardslee, D. S., & Wertheimer, M. (*Eds.*). Readings in Perception. Princeton: Nostrand, 1958. Pp. 751.
Bond, E. D. One Mind, Common to All. New York: Macmillan, 1958. Pp. 200.
Edwards, A. L. Statistical Analysis. New York: Rinehart, 1958. Pp. 234.
Fisher, S., & Cleveland, S. E. Body Image and Personality. Princeton: Nostrand, 1958. Pp. 420.
Fraisse, P. Psychologie du Temps. Paris: Presses Universitaires de France, 1957. Pp. 326
Gardner, D. E. M. The Education of Young Children. New York: Philosophical Library, 1957. Pp. 118.
Gardner, M. Fads and Fallacies in the Name of Science. New York: Dover, 1957. Pp. 363.
Garrett, E. J. (*Ed.*). Beyond the Five Senses. Philadelphia: Lippincott, 1957. Pp. 384.
Gibson, W. The Miracle Worker. New York: Knopf, 1957. Pp. 131.
Graham, T. F. Dynamic Psychopathology. Boston: Christopher, 1957. Pp. 251.
Gebhard, P. H., Pomeroy, W. B., Martin, C. E., & Christenson, C. V. Pregnancy. Birth and Abortion. New York: Harper, 1958. Pp. 282.
Greenblatt, M., Levinson, D. J., & Williams, R. H. (*Eds.*). The patient and the mental hospital. Glencoe, Ill.: Free Press, 1957. Pp. 658.
Halmos, P. Towards a Measure of Man. New York: Humanities, 1957. Pp. 250.
Hatt, P. K., & Reiss, A. J., Jr. (*Eds.*). Cities and Society. (2nd Ed.) Glencoe: Free Press, 1957. Pp. 852.
Heath, S. Citadel, Market and Altar: Emerging Society. Elkridge, Md.: Science of Society Foundation, 1957. Pp. 259.
Hoult, T. F. The Sociology of Religion. New York: Dryden, 1958. Pp. 436.
James, E. O. Prehistoric Religion: A Study in Prehistoric Archaeology. New York: Praeger, 1957. Pp. 300.
Jones, E. The Life and Work of Sigmund Freud. Vol. III. New York: Basic Books, 1957. Pp. 537.
Jones, M. R. (*Ed.*). Nebraska symposium on motivation, 1957. Lincoln: Univ. Nebraska Press, 1957. Pp. 430.
Konopka, G. Eduard C. Lindeman and Social Work Philosophy. Minneapolis: Univ. Minnesota Press, 1958. Pp. 220 ,
Krugman, M. (*Ed.*). Orthopsychiatry and the School. New York: American Orthopsychiatric Assoc., 1958. Pp. 265.

LEAVITT, H. J. Managerial Psychology. Chicago: Univ. Chicago Press, 1958. Pp. 334.

LUCE, H. D., & RAIFFA, H. Games and Decisions: Introduction and Critical Survey. New York: Wiley, 1957. Pp. 509.

MACINTYRE, A. C. The Unconscious. New York: Humanities Press, 1958. Pp. 100.

MARTINEAU, P. Motivation in Advertising: Motives that Make People Buy. New York: McGraw-Hill, 1957. Pp. 210.

MCCARTHY, P. J. Introduction to Statistical Reasoning. New York: McGraw-Hill, 1957. Pp. 402.

MCGUIGAN, F. J., & CAVLIN, A. D. Current Studies in Psychology. New York: Appleton-Century-Crofts, 1958. Pp. 226.

MERTON, R. K., READER, G., & KENDALL, P. L. (*Eds.*). The Student-Physician: Introductory Studies in the Sociology of Medical Education. Cambridge: Harvard Univ. Press, 1957. Pp. 360.

MICHAELS, J. J. Disorders of Character: Persistent Enuresis, Juvenile Delinquency, Psychopathic Personality. Springfield: Thomas, 1955. Pp. 148.

MOLONEY, J. C. Fear: Contagion and Conquest. New York: Philosophical, 1957. Pp. 140.

MUDD, E. H., & KRICH, A. (*Eds.*). Man and Wife. New York: Norton, 1957. Pp. 291.

MUELLER, K. H. Educating Women for a Changing World. Minneapolis: Univ. Minnesota Press, 1954. Pp. 302.

MULLANY, P. (*Ed.*). A Study of Interpersonal Relations. New York: Grove, 1957. Pp. 507.

NADLER, G. Work Simplification. New York: McGraw-Hill, 1957. Pp. 292.

NEUSTATTER, W. L. Psychological Disorder and Crime. New York: Philosophical, 1957. Pp. 248.

NOLL, V. H. Introduction to Educational Measurement. Boston: Houghton-Mifflin, 1957. Pp. 437.

OATES, W. E. Religious Dimensions of Personality. New York: Association Press, 1957. Pp. 320.

PAGE, R. C. It Pays to be Healthy. Englewood Cliffs, N. J.: Prentice-Hall, 1957. Pp. 285.

PALMER, H. Psychopathic Personalities. New York: Philosophical, 1957. Pp. 179.

PETERS, R. S. The Concept of Motivation. New York: Humanities, 1958. Pp. 166.

PHELPS, H. A., & HENDERSON, D. Population in its Human Aspects. New York: Appleton-Century-Crofts, 1958. Pp. 512.

PICKER, L. My Inward Journey. Philadelphia: Westminster, 1957. Pp. 187.

PIOTROWSKI, Z. A. Perceptanalysis. New York: Macmillan, 1957. Pp. 505.

POLYAK, S. The Vertebrate Visual System. Chicago: Univ. Chicago Press, 1957. Pp. 1390.

PRESCOTT, D. A. The Child in the Educative Process. New York: McGraw-Hill, 1957. Pp. 502.

PRINCE, M. The Dissociation of a Personality. New York: Longmans, Green, 1957. Pp. 575.

RADIN, P. Primitive Religion. New York: Dover, 1957. Pp. 232.

REDL, F., & WINEMAN, D. The Aggressive Child. Glencoe: Free Press, 1957. Pp. 575.

REIK, T. Of Love and Lust. New York: Farrar, Straus & Cudahy, 1957. Pp. 623.

RHINE, J. B., & PRATT, J. G. Parapsychology. Springfield: Thomas, 1957. Pp. 220.

ROSS, M. G., & HENDRY, C. D. New Understandings of Leadership. New York: Association Press, 1957. Pp. 158.

ROBINSON, E. G., JR. My Father—My Son. New York: Fell, 1958. Pp. 316.

RUCH, F. L. Psychology and Life. Chicago: Scott, Foresman, 1958. Pp. 624.

SAMSTAG, N. Persuasion for Profit. Norman: Univ. Oklahoma Press, 1957. Pp. 208.

SARGENT, S. S., & WILLIAMSON, R. C. Social Psychology. New York: Ronald Press, 1958. Pp. 649.

SARTAIN, A. Q., NORTH, A. J., STRANGE, J. R., & CHAPMAN, H. M. Psychology: Understanding Human Behavior. New York: McGraw-Hill, 1958. Pp. 444.

SCHAFFNER, B. (*Ed.*). Group Processes. New York: Josiah Macy, Jr., 1957. Pp. 328.
SCHMEIDLER, G. R., & MCCONNELL, R. A. ESP and Personality Patterns. New Haven: Yale Univ. Press, 1958. Pp. 136.
SCHREIER, F. T. Human Motivation. Glencoe: Free Press, 1957. Pp. 256.
SELENICK, P. Leadership in Administration. Evanston: Row, Peterson, 1957. Pp. 162.
SHOCK, N. W. Trends in Corontology. Stanford: Stanford Univ. Press, 1957. Pp. 214.
SOROKIN, P. Social and Cultural Dynamics. Boston: Porter Sargent, 1957. Pp. 718.
SOWER, C., HOLLAND, J., TIEDKE, K., & FREEMAN, W. Community Involvement. Glencoe: Free Press, 1957. Pp. 323.
STOCK, D., & THELEN, H. A. Emotional Dynamics and Group Culture. New York: New York Univ. Press, 1958. Pp. 296.
STRANG, R. Group Work in Education. New York: Harper, 1958. Pp. 322.
TATE, M. W., & CLELLAND, R. C. Non-parametric and shortcut statistics. Danville, Ill.: Interstate, 1957. Pp. 171.
THORPE, L. P., & SCHMULLER, A. M. Personality. Princeton: Van Nostrand, 1958. Pp. 368.
WALKER, H.M., & LEV, J. Elementary Statistical Methods. New York: Holt, 1958. Pp. 302.
WAPNER, S., & WERNER, H. Perceptual Development. Worcester: Clark Univ. Press, 1957. Pp. 95.
WEITZENHOFFER, A. M. General Techniques of Hypnotism. New York: Grune & Stratton, 1957. Pp. 460.
WINNICOTT, D. W. Mother and Child. New York: Basic Books, 1957. Pp. 210.

(Packard, V. The Hidden Persuaders. New York: McKay, 1957. Pp. 275.)

REVIEWED BY HENRY WINTHROP

In recent years there has been a steady stream of material on the manner in which psychological findings and psychological techniques are being used to usher in a "brave, new world." We have had studies of brain-washing, spell-binding, the misuse of tests by organization men and the engineering of consent, as examples of Exhibit *A*. The volume being reviewed, which is devoted to an examination of the use of psychological techniques in creating illusory wants for consumers, is one of the more recent contributions in this direction.

The area of *motivation research* has been riding high for some time now and this volume constitutes an effort to describe its alleged triumphs and failures. Packard is a newspaper man for whom description and evaluation take precedence over analysis. This, I suppose, is inevitable, considering that the volume was intended for an intelligent lay audience, although as a running summary of and commentary on the work in the field of motivation research, it will also prove of value to those professional psychologists who have little or no acquaintance with the subject and who can use an enlightened account free of professional technicalities. Packard leans heavily upon Cheskin's definition of motivation research, which runs as follows: "Motivation research is the type of research that seeks to learn what motivates people in making choices. It employs techniques designed to reach the unconscious or subconscious mind because preferences generally are determined by factors of which the individual is not conscious. . . . Actually in the buying situation the consumer generally acts emotionally and compulsively, unconsciously reacting to the images and designs which in the subconscious are associated with the product."

Packard divides his volume into four parts bearing the titles: The Depth Approach, Persuading Us As Consumers, Persuading Us As Citizens, and In Retrospect. Most of the significant material is in the second portion. The variety of consumer motivation which the symbol manipulators are unearthing can be gathered from some of the findings the author reports. Among these are the following. First there is Dr. Dichter's famous study "Mistress versus Wife" in which the male consumer was persuaded to buy convertibles rather than sedans by introducing the hardtop convertible whose symbolic value to the male buyer was that it was supposed to suggest the lure of a mistress together with the solid fidelity and down-to-earth qualities of the plain girl he married, who became a good wife and mother. Then there is the depth study which put the cigar industry on its feet by discovering that cigars are virility symbols and that by smoking them, particularly in the presence of feminine advocates of gracious living, the male is defiantly reasserting his masculinity. Another intriguing study reported by Packard was the detection of the state of hypnosis in which the female shopper finds herself in the average supermarket. The state is induced by the packaged fairyland before her and it is detected by the subnormal eye-blink rate which accompanies it. The researcher who discovered this also unearthed the startling fact that feminine resistance to the early cake mixes to which one had only to add water and bake, lay in the *fact* that "baking a cake traditionally is acting out the birth of a child" and that having to do nothing to get a cake is like being deprived of one's reproductive capacities and feminine status, and no proper lady wants a culinary pseudocyesis. The motivational researchers have also found that the most successful cereals for children are those with a built-in *crunch*. This appeases hostility by supplying an outlet for aggression and other feelings. It is unpleasant to have to report this but motivation researchers have also found that men's clothing styles are becoming feminized because of "creeping momism." Women buy their husbands' clothing to improve their "public images" and the current generation of momistically, other-oriented males actually want their wives to be mother-surrogates in this connection. Last but not least of the motivation research studies worth mentioning here is the discovery that automobiles are heavily laden with social meanings which "provide avenues for the expression . . . of the character, temperament, and self concept of the owner and driver. . . . The buying process is an interaction between the personality of the car and the personality of the individual." People who want to seem sociable and up-to-date in a middle-of-the-road sort of way buy Chevrolets, Pontiacs, Buicks and Chryslers. People who want to be showy and assert their individualism and modernity purchase a Ford, Mercury, Oldsmobile, or Lincoln. People who need to express unusual status or individual needs favor the Cadillac, Studebaker, Hudson, Nash, and Willys.

Packard devotes the material of Parts II and III to discussing the activities of symbol manipulators in engineering consent, promoting alienation, producing a sense of social marginality and in promoting what Fromm has called the marketing orientation. The Orwellian possibilities in motivation research disturb Packard. He quotes material from an article entitled "Custom-made Men" which appeared in *Diesel Power* to show the lengths to which motivation researchers are prepared to go. He points out how advertising men with the advice of "leading child psychologists and psychiatrists," using projective and sentence-completion tests, broke down maternal resistance to giving home permanents to their little daughters, by showing that the little daughters equated loneliness with having straight hair and that having straight hair was equated with "bad, unloved things." One gets a decidedly uncomfortable feeling as Packard sets forth the many socially questionable uses of motivation research and many a reader will rush off to take a fresh look at Walden II after finishing Packard's volume.

The anecdotes furnished by Packard are impressive—if you like to be impressed by being depressed. However, it is to be noted that the scientific basis of many of the studies quoted leave much to be desired. The language of depth research is applied by Madison Avenue in even murkier ways than it is used by its initiators, and in some respects the motivation researchers in Madison Avenue rookeries dredge deeper and come up with even less. Field investigations in motivation research frequently make use of incautious experimental designs. Interpretations of interviewees' comments are often riotously imaginative and validated only by the use of projective tests. The use made of such projective tests is amply demonstrated by the following comment, made in connection with the problem of trying to sell life insurance to men. ". . . projective tests revealed the respondent's fierce desire to achieve immortality in order to *control* his family after death. These men obtain insurance against obliteration through the knowledge that they will continue to *dominate* their families; to *control* the family standard of living, and to guide the education of their children long after they are gone." Advertising campaign decisions based upon motivation research are frequently rooted in highly colored interpretations of word-association tests. Finally the confirmation of advice based upon psychological findings is often only a sales spurt with no check made to see if this spurt is a typical sales fluctuation or not and no effort made to see whether such a spurt is a *recency effect* of a sales campaign rather than a persuasion which will not stay put. The major complaints which may be brought against motivation research may then be summarized as follows: (*a*) there is no single or major reason why people buy or do not buy a product; (*b*) the clinical tools employed and applied to mass behavior have not been validated in this new context; (*c*) standardizing and validating procedures for methods other than the clinical devices employed, are conspicuous by their absence and (*d*) initial hunches in motivation research are often also the final conclusions with little effort to confirm them through the establishment of experimental controls.

Packard also wrestles with the morality of motivation research and needless to say morality comes out second best. Regardless, however, of the reader's social philosophy, psychologists will be interested in acquainting themselves with this new area of applied psychology, which promises to grow larger before it grows more suspect. Motivation research may yet prove to be an additional chapter in the next edition of Eysenck's *The Uses and Abuses of Psychology*.

Department of Psychology
University of Wichita
Wichita 14, Kansas